THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE

Beyond western, educated, industrial, rich, and democratic (WEIRD) psychology: measuring and mapping scales of cultural and psychological distance

Article:

Beyond WEIRD Psychology: Measuring and Mapping Scales of Cultural and Psychological

Distance

Michael Muthukrishna[1], Adrian V. Bell[2], Joseph Henrich[3], Cameron M. Curtin[3], Alexander

Gedranovich[1], Jason McInerney[4], Braden Thue[3]

[1] London School of Economics and Political Science

[2] University of Utah

[3] Harvard University

[4] Iowa State University

Abstract

We present a tool and method for measuring the psychological and cultural distance between societies and creating a distance scale with any population as the point of comparison. Since psychological data is dominated by samples drawn from WEIRD nations, and overwhelmingly, the United States, we focus on distance from the US. We also present distance from China, the largest population, second largest economy, and common cultural comparison. We apply the fixation index (F_{ST}), a meaningful statistic in evolutionary theory, to the World Values Survey of cultural beliefs and behaviors. As the extreme WEIRDness of the literature begins to dissolve, the tool will become more useful for designing, planning, and justifying a wide range of comparative psychological projects. Our code and accompanying online application allow for comparisons between any two countries. Analyses of regional diversity reveal the relative homogeneity of the United States. Cultural distance predicts various psychological outcomes.

Keywords: WEIRD people; cultural psychology; cultural distance; cross-cultural differences; replication crisis

Beyond WEIRD Psychology: Measuring and Mapping Scales of Cultural and Psychological

Distance

Decades of psychological research designed to uncover truths about human psychology

may have instead uncovered truths about a thin slice of our species – those who live in Western

Educated Industrialized Rich Democratic (WEIRD) nations (Henrich, Heine, & Norenzayan,

2010). Researchers often assess the generalizability of these findings by comparing Western

nations to East Asian nations, but are increasingly documenting differences in small-scale

societies. Nonetheless, the literature remains overwhelmingly WEIRD (Rad, Martingano, &

Ginges, 2018) and there exists no systematic method for determining which societies will

provide useful comparisons or even the size of the psychological differences—the cultural

distance—between societies, be they non-Western, less-educated, less-industrialized, poorer,

non-democratic or some subset of these. And even within WEIRD nations, there are

psychologically-relevant cultural differences (Henrich et al., 2010; McCrae, Terracciano, & 79

Members of the Personality Profiles of Cultures Project, 2005). A growing body of theoretical

and empirical work in cultural evolution emphasizes that our species is fundamentally cultural,

and thus these cultural differences are also psychological differences: from norms and attitudes,

to the degree to which these norms are enforced, to low-level perception of color and visual

illusions (Boyd, 2017; Gelfand, 2019; Henrich, 2016).

Just how psychologically different are the nations of the world compared to each other

and to the over-scrutinized United States? Many hard drives have been filled with the ways in

which China and Japan differ from the United States and Canada, but just how psychologically

distant is the culture of China *from* Japan, the United States from Canada, or Azerbaijan from

Zambia? Here we introduce a robust method for quantifying this distance. This method allows us

to develop scales of cultural distance, and therefore cross-cultural psychological distance[1], by selecting any population as a point of comparison. Since psychological data remains largely American (Rad et al., 2018), as an example we develop an *American scale* of cultural distance from the United States. As a point of comparison, we also develop a *Sino scale* of cultural distance from China, the largest population on Earth and a common cultural comparison. Using our *R* code or online tool (www.culturaldistance.com), researchers can create scales of cultural distance with any comparison population, which will become increasingly important as the literature becomes less WEIRD.

The measurement of cross-cultural psychological differences and cultural distances has a long history. Apart from the many differences identified by cultural psychologists (Heine, 2015; Henrich et al., 2010), notable attempts to quantify these differences include Hofstede's Cultural Dimensions (Hofstede, 2001), Ingelhardt and Welzel's cultural map (Inglehart & Welzel, 2005), and Schwartz's values (Schwartz, 2006). These difference measures are sometimes combined and used as distance measures. Notable examples include Kogut and Singh's (1988) composite measure of Hofstede's Cultural Dimensions, and Demes and Geeraert's (2014) scale of perceived cultural distance. Within economics and political science, genetic distance and linguistic distance are often used as proxies for cultural distance (Desmet, Ortuño-Ortín, & Wacziarg, 2017; Gorodnichenko & Roland, 2017; Spolaore & Wacziarg, 2016).

These approaches are widely used, but have various limitations. For example, the values and dimensions approach characteristic of Hofstede, Ingelhardt, and Schwartz focuses on identifying the values or dimensions along which groups differ, focusing on cultural *differences*

---

[1] By cross-cultural psychological distance, we refer to the size of the difference in psychology between different societies rather than the perceived cognitive distance between the self and other individuals, objects or events (psychological distance in construal level theory).

rather than cultural *distance*. Moreover, these values represent mean differences, largely ignoring differences in variance or frequencies of beliefs and behaviors – bimodal or multimodal populations appear the same if they have the same mean. For example, Brazil and Turkey have almost identical scores along Hofstede's Individualism dimension (38 and 37). Indeed, Brazil and Turkey look very similar along most Hofstede dimensions (see Figure 1). Brazil and Turkey differ in a variety of ways, however, overall Brazil has greater variability in beliefs and values than Turkey (variance in beliefs is sometimes measured as "looseness" or tolerance for deviant behavior; Brazil and Turkey are on opposite ends of this spectrum; Gelfand et al., 2011). But looseness too is a point estimate – variance in cultural traits vary by domain, even if some nations tend to be tighter or looser overall. And when it comes to nominal cultural traits, such as whether a participant's political priorities are to "give people more to say", "maintain order in the nation", "fight rising prices", or "protect freedom of speech", neither mean nor variance capture the relative frequencies.
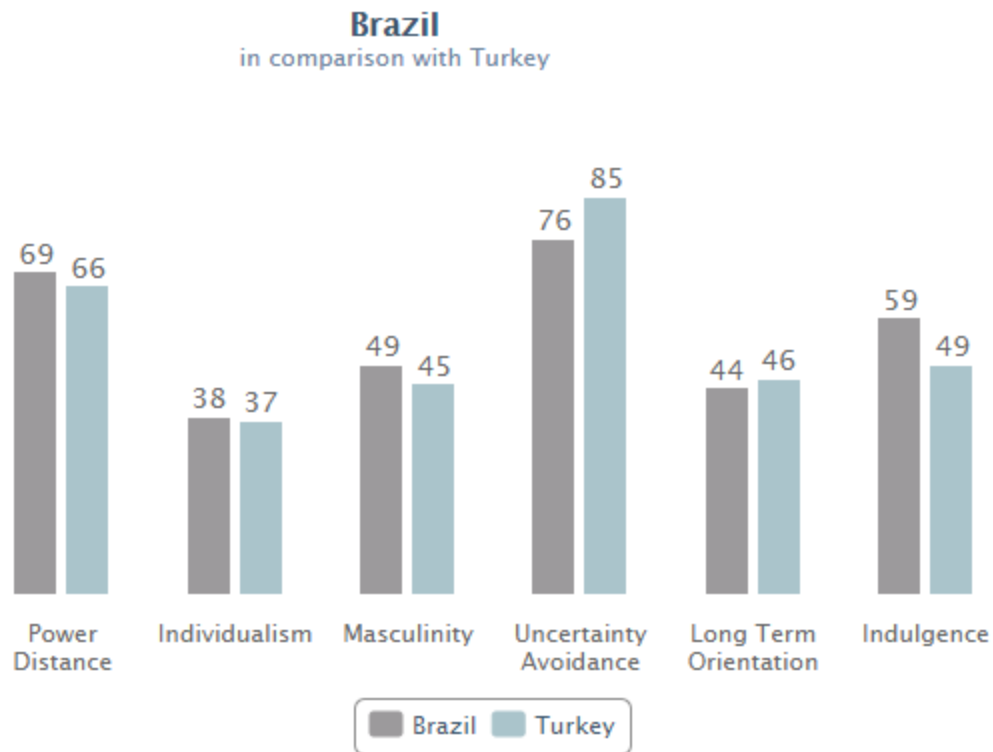
Figure 1. Brazil vs. Turkey on Hofstede Cultural Dimensions (Source: geert-hofstede.com).

Here we present Cultural F_{ST} (CF_{ST}), a measure built on the F_{ST} measure from population biology (Bell, Richerson, & McElreath, 2009; Cavalli-Sforza, Menozzi, & Piazza, 1994), as a tool for measuring cultural distance. CF_{ST} is robust and theoretically defensible, and can be used as a high-resolution method to identify regional, national, or arbitrary cultural groupings (such as class differences). It can be used to identify how distant two groups are based on an aggregate of many cultural dimensions or along any arbitrary dimensions, such as politics or social relations, depending on the theory being tested (Muthukrishna & Henrich, 2019). We first discuss the technique and then use it to construct example scales of cultural distance that can guide researchers in deciding where to target their data collection efforts. To make it easy for researchers to use this method, we have made the *R* code available and developed an online tool (http://culturaldistance.com). The online tool allows researchers to explore particular dimensions

of difference and future versions of the tool will allow comparisons at the regional level. A guide

to using the online tool is available in the Supplementary Materials.

F_{ST} is theoretically meaningful within evolution, because it measures how genotype

frequencies for each subpopulation differ from expectations if there were random mating over

the entire population; that is, it measures the degree to which the populations can be considered

structured and separate. For cultural inheritance, this is directly analogous to between-group

differentiation caused by selection, migration, and social learning mechanisms. Thus, insofar as

one advocates a formal theoretical approach to psychological research (Muthukrishna & Henrich,

2019) and insofar as cultural evolution offers a compelling explanation for cultural change, F_{ST}

offers a theoretically informed approach to measuring cultural distance with a long and

established history within the broader evolutionary sciences. We discuss this in more detail,

including comparisons to other methods in the Supplementary Material.

We begin by comparing CF_{ST} to other common approaches to measuring cultural

distance. We then formally describe the CF_{ST} statistic and use it to develop the American and

Sino scales. We demonstrate the robustness of the scales to missing or incomplete data and show

how it can also be used to study cultural variation within a population, using regional variation as

an example. Finally, we show the relationship between these scales and other measures of

cultural distance and also test how well they predict common or high-profile measures of cross-

cultural psychological differences.

Methods

One approach to measuring cultural distance is to turn difference measures, such as

Hofstede's Cultural Dimensions, into a distance measure by taking a weighted mean of the

distance between a country's values on the difference scale. The most popular of these composite

measures is the formulation by Kogut and Singh (1988):

$$CD_{cj} = \sum_{i=1}^{d} \left(\left(I_{ij} - I_{ic}\right)^2 \Big/ V_i \right) \Big/ d$$

Where cultural distance CD between country c and country j is the mean over all

dimensions d of the squared difference between the countries' cultural difference value I for

dimension i weighted by the variance V of that dimension i. This method allows researchers to

turn these differences into a composite distance measure that captures a distance between

countries on Hofstede's cultural dimensions. However, it suffers the same limitations as the

underlying difference measures, such as focusing on mean differences between countries and

ignoring differences in variance or frequencies (the variance above is the variance of the index

across countries, not variance in culture within the country). Brazil and Turkey still look

culturally close because they have similar mean cultural difference values along most

dimensions.

Another approach, as exemplified by Demes and Geeraert (2014) is to develop a scale

that asks participants to compare countries on perceived cultural distance along dimensions such

as climate, food and eating, values and beliefs, and so on. This is similar to the development of

any psychometric scale, but is a measure of the perception of participants from Country A about

Country B. This approach may be a good measure of perceptions of culture, but may be limited

by participants of Country A's knowledge of participants from Country B. Here, we will use the

World Values Survey as a measure of what people themselves report they do and believe.

Genetic distance and linguistic distance do not share these limitations, serving as an

overall proxy measure of cultural distance between any two nations. But by not measuring

culture directly, they can be misleading. For example, Hong Kong is more than an order of magnitude more genetically similar to China ($F_{ST} = 9.59 \times 10^{-4}$) than to Great Britain ($F_{ST} = 3.96 \times 10^{-2}$) (Pemberton, DeGiorgio, & Rosenberg, 2013; Spolaore & Wacziarg, 2016), but is culturally similar to both due to Britain's century-long history in Hong Kong. Linguistic distance is a better measure of cultural distance—language is a core aspect of culture and the lack of a common language can culturally isolate groups—but the resolution of language distance makes it difficult to distinguish the cultures of Australia, Canada, the UK and the US, all of whom speak English. These differences drive tourism and migration, but are largely invisible on a language family tree. None of these measures serve as a robust and defensible cultural distance scale that can be used to compare two sampled populations, an important tool given the stark psychological differences between cultural groups (Henrich et al., 2010).

Fixation Index (F_{ST})

Fixation index or F_{ST} (Cavalli-Sforza et al., 1994) is a method used to calculate the size of the genetic differences (genetic distance) between two populations. In the genetic case, F_{ST} represents the ratio of the between- and within-group variance of alleles (such as gene variants for blue or brown eyes) at a particular locus (such as the DNA location for the main eye color gene) in the genomes of individuals in two populations. Formally:

$$F_{ST} = \frac{\sigma_g^2}{\sigma_T^2}$$

Where σ_g^2 is the variance of allele frequency between populations and σ_T^2 is the variance of allele frequency in the total population. In practice, F_{ST} is calculated as:

$$F_{ST} = \frac{H_T - H_g}{H_T}$$

Where H_T is the average number of pairwise differences in alleles (for genetic distance) or answers (for cultural distance) between individuals drawn from the total population (both populations) and H_g is the average number of pairwise differences in alleles or answers between two individuals from the same population. We can see that if these pairwise differences are the same, $F_{ST} = 0$ – the populations are identical. If the two equal sized populations are more homogenous, but different, we get the maximum distance of $F_{ST} = 1$ (see Figure 2). Like a correlational coefficient, F_{ST} is also a non-linear metric. This means that the numbers are meaningful for comparisons, but $F_{ST} = 0.6$ is not twice as large as $F_{ST} = 0.3$ (just as a correlation of $r = 0.6$ is not twice as high a correlation at $r = 0.3$)
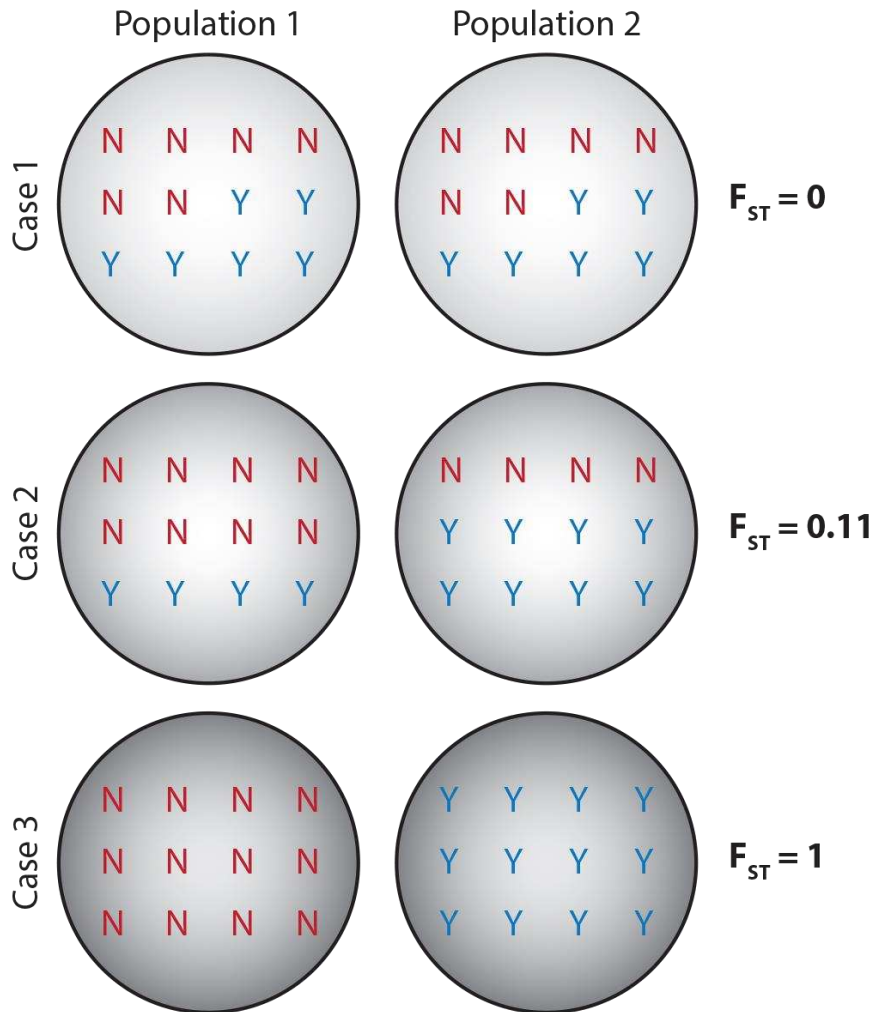
Figure 2. F_{ST} calculated along a single dimension with two options – Yes (Y) or No (N). In Case 1, the cultural distance between the two populations is 0 since 50% of both populations answer Yes (or No). In Case 2, F_{ST} is 0.11. Finally, in Case 3, the populations are maximally different, since all individuals in Population 1 say No and all individuals in Population 2 say Yes. Calculations for each of these cases can be found in the Supplementary Material. For two populations, we calculate mean F_{ST} along all cultural questions or along specific questions of interest (such as those mapping onto a particular dimension).

To get an overall genetic distance between two populations, one can aggregate the F_{ST} values for allele frequencies across all loci or some subset loci of interest in the genome of a representative sample of each population. The availability of large representative cross-national surveys of cultural values has allowed the same technique to be applied to culture.

Summary of Cultural F_{ST} (CF_{ST})

Cultural F_{ST} (CF_{ST}) is calculated in the same manner as Genetic F_{ST}, but instead of a genome, we use a large survey of cultural values as a "culturome", with questions treated as loci and answers treated as alleles. Note, strictly speaking, the answers to such questions are phenotypes – that is, responses are based on a combination of culture, genes, personal experience, and perhaps other inheritance systems (epigenetics, microbiota, etc.; Laland, 2017). This does not affect use for scales of cultural distance and, arguably, what we are really interested in are these expressed phenotypic differences. CF_{ST} allows us to flexibly quantify the overall size of cultural differences (cultural distance). In contrast to many other measures of cultural differences, CF_{ST} compares distributions rather than point estimates (not simply that one group has a higher mean than another), does not assume homogeneity in groups (instead looking at the frequencies of cultural traits), and does not assume that traits fall along a single dimension (instead individuals can be higher or lower along related continuous cultural traits or have a range of discrete, orthogonal cultural traits). Moreover, CF_{ST} can handle continuous, binary, or nominal traits. By calculating CF_{ST} for subsets of questions, we can also see how nations differ along different dimensions. For example, a "family values" dimension might include questions on the "importance of family", "respect for parents", "parents' duty towards their children", and the various values parents wish to instill in their children. These dimensions may be pre-determined (similar to including all genes associated with height or skin color) or based on dimension reduction techniques. A list of pre-determined dimensions and statistically reduced dimensions can be found in the section on Data. You can examine pre-determined dimensions using http://culturaldistance.com. Of course, the use of particular dimensions should be theoretically motivated (see Muthukrishna & Henrich, 2019), and so here we use an aggregate

measure to demonstrate the use of CF_{ST}. CF_{ST} can also be calculated between regions within larger nations; nations are not equally heterogeneous. Such within-country variation is important for assessing generalizability.

Using this common currency, we can apply several visualization techniques commonly used in bioinformatics, such as neighbor-joining to cluster countries that are most similar, multidimensional scaling plots to visualize diversity within a nation or identify "cultural continents", and density plots to investigate subdimensions within dimensions or questions within subdimensions (examples in Supplementary Materials). Moreover, we can calculate confidence intervals as a measure of uncertainty.

Formal Description of Cultural F_{ST} (CF_{ST})

Cultural F_{ST} (CF_{ST}) can be computed for any measured trait as long as there is data measured at an individual level and we know the group or groups to which the individual belongs. The data can be continuous (quantitative, cardinal, ordinal data) or categorical (nominal). CF_{ST} is computed for a particular trait. To calculate the CF_{ST} between groups, we can use the mean CF_{ST} across all traits for an overall distance measure or all traits of interest for a particular domain or dimension. When measuring genetic distance, one could measure overall genetic distance or only the distance for loci associated with a particular disease or associated with a physical trait, such as height. Similarly, for CF_{ST}, we might only be interested in traits associated with prosociality, sexual attitudes or political positions. Focusing on specific dimensions may be important for particular research questions, driven by a particular theory (Muthukrishna & Henrich, 2019), but since we have no particular theoretical basis for examining subdimensions, here we will focus on overall cultural distance as a demonstration. Let's begin by

explaining the CF_{ST} formula, starting with continuous data, which is more straightforward and should be more familiar.

Continuous data

As discussed, at its core, CF_{ST} is the ratio of the between-group variance (σ_g^2) and total variance (σ_T^2):

$$CF_{ST} = \frac{\sigma_g^2}{\sigma_T^2}$$

We can use standard ways of computing a statistical variance when dealing with quantitative characters and in some cases cardinal and ordinal data. The variance for the whole population is taken by summing across all individuals in the population:

$$\sigma_T^2 = \frac{\sum_{i=1}^{s}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2}{N - 1}$$

where x_{ij} is the quantitative measure of an observation from individual j in group i, n_i is the number of individuals in group i (with sample size s), $\bar{x}$ is the mean trait value summed across all individuals in the population, and $N = \sum_i n_i$.

Similarly, the variance between groups σ_g^2, can be computed as:

$$\sigma_g^2 = \frac{\sum_{i=1}^{s}(x_i - \bar{x})^2}{s - 1}$$

where x_i is the mean quantitative trait value in group i.

Categorical data

For categorical data, we can adapt the Equation 4 formula in Table 1.10.1 of Cavalli-Sforza et al. (1994). For a question or loci with L number of outcomes and p_{ki} as the frequency of outcome (e.g. answer) k in group i, we can compute the CF_{ST} for a particular observation:

$$CF_{ST,k} = \frac{var(p_k)}{\bar{p}_k(1 - \bar{p}_k)}$$

where

$$\bar{p}_k = \frac{\sum_{i=1}^{S} n_i p_{ki}}{\sum_{i=1}^{S} n_i}$$

is the average allele frequency across s populations weighted by sample size (n_i) and

$$var(p_k) = \frac{\sum_{i=1}^{S} (p_{ki} - \bar{p}_k)^2}{s - 1}$$

is the between group variance in observed frequencies of answers.

This specification gives us the deviations from the mean trait frequency across all groups. Across all questions the CF_{ST} is:

$$CF_{ST} = \frac{\sum_{k=1}^{L} \bar{p}_k(1 - \bar{p}_k)CF_{ST,k}}{\sum_{k=1}^{L} \bar{p}_k(1 - \bar{p}_k)}$$

Data

Cultural distance is calculated using combined data from the two most recent waves of the World Values Survey (WVS; 2005-2009 & 2010-2014). These waves contain answers from 170,247 participants gathered from nationally representative samples of 80 countries, where approximately 85% of the world live. We included values, beliefs, and behaviors that we judged as culturally transmissible, largely similar to those selected by Bell et al. (2009), but excluded questions that were specific to a region (e.g. confidence in the North American Free Trade Agreement; NAFTA) – Bell and colleagues only compared neighboring countries within the same region so were able to include these. We also exclude all demographic questions (e.g. age

and sex). A full list of our inclusion-exclusion decisions for all WVS questions is available in the

Supplementary Materials.

We use CF_{ST} to develop an American scale of cultural distance from the United States

and a Sino scale of cultural distance from China. We compare these scales to the following

cultural differences, psychological outcomes, and distance measures.

Hofstede's Cultural Dimensions

The Hofstede cultural dimensions (Hofstede, 2001) were originally based on surveys of

IBM staff collected between 1967 and 1973. The original dimensions included *Collectivism-*

Individualism, *Power Distance*, *Femininity-Masculinity*, and *Uncertainty Avoidance*. Two

additional dimensions, *Long-term Orientation* and *Indulgence* were added later.

Hofstede offers the following descriptions on his website: Individualism is the extent to

which people feel independent, as opposed to being interdependent as members of larger wholes.

Power Distance is the extent to which the less powerful members of organizations and

institutions (like the family) accept and expect that power is distributed unequally. Masculinity is

the extent to which the use of force in endorsed socially. Uncertainty avoidance deals with a

society's tolerance for uncertainty and ambiguity. Long-term orientation deals with change. In a

long-time-oriented culture, the basic notion about the world is that it is in flux, and preparing for

the future is always needed. In a short-time-oriented culture, the world is essentially as it was

created, so that the past provides a moral compass, and adhering to it is morally good.

Indulgence is about the good things in life. In an indulgent culture it is good to be free. Doing

what your impulses want you to do is good. Friends are important and life makes sense. In a

restrained culture, the feeling is that life is hard, and duty, not freedom, is the normal state of

being.

Tightness-looseness

Gelfand et al. (2011) describes tight societies as having many strong norms and a low tolerance of deviant behavior. In contrast, loose societies are characterized as having weak social norms and a high tolerance of deviant behavior. Gelfand et al. (2011) measures tightness based on perceptions of social norms and norm enforcement using a survey that includes questions such as "There are many social norms that people are supposed to abide by in this country." and "In this country, if someone acts in an inappropriate way, others will strongly disapprove.".

More recently, Uz (2015) has suggested measuring tightness-looseness based on its outcome—greater variance or standard deviation in professed cultural values. Uz (2015) introduced three SD measures of answers in the 2000 wave of the World Values Survey. These included (1) a domain-specific index of standard deviation of answers to the Morally Debatable Behaviors Scale, which included attitudes towards prostitution, abortion, divorce, euthanasia, and suicide; (2) a domain-general index based on the standard deviation of a variety of values and behavioral practices in the World Values Survey; and (3) a composite measure based on a factor analysis of the domains in the World Values Survey. This combined or composite measure had the greatest validity. All Uz (2015) measures of looseness are only weakly correlated with the Gelfand et al. (2011) measure ($r = .16$ to $.30$).

Schwartz's Values

Schwartz's Cultural Value orientation is a theory that outlines various values that help regulate human behavior in different societies (Schwartz, 2006). Harmony refers to fitting into the world as it is, whereas mastery refers to the tendency to change the world to achieve your goals. Affective autonomy refers to pursuing pleasurable experiences. Intellectual autonomy refers to pursuing your own ideas independently. Embeddedness is similar to collectivism.

Hierarchy refers to unequal distribution of power, whereas egalitarianism refers to recognizing all people as moral equals.

Five Factor Model of Personality

We use cross-cultural data gathered by McCrae et al. (2005) for the five factor model of personality. These personality factors include Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These data were gathered with between 133 and 919 participants within each of 50 societies using the 240-item Revised NEO Personality Inventory (NEO-PI-R) measure given in the participants' native language. The Standard Deviation was the mean T-score-standardized standard deviation across 30 NEO-PI-R facet scales. Recent research suggests that variance in personality is predicted by sociocultural complexity (Smaldino, Lukaszewski, von Rueden, & Gurven, 2019).

Other Psychological and Behavioral Measures

We also included several other measures that have been used by psychologists, behavioral scientists and economists. With the exception of the corruption perceptions index, these variables capture psychological outcomes that have been suggested to have WEIRD origins (Schulz, Bahrami-Rad, Beauchamp, & Henrich, 2019).

Blood donations per 1000 individuals was used as a measure of impersonal cooperative altruism. It measures the frequency of voluntary, unpaid, anonymous blood donations from countries (WHO) and Italian provinces. This data was collated by (Schulz et al., 2019) from the WHO Global Status on Blood Safety and Availability. Diplomat Parking Tickets is based on data on unpaid parking tickets of United Nations diplomats in New York City. This data comes from a classic economics study showing the number of unpaid parking tickets are predicted by corruption norms (Fisman & Miguel, 2007). Diplomats do not face sanctions for unpaid parking

tickets due to diplomatic immunity, but the rate of accruing tickets varies considerably by country of origin, ranging from a mean of 0 per diplomat to 249 per diplomat. The corruption perceptions index is a measure of the descriptive corruption norm published by Transparency International, using the 2015 index to match the WVS. In addition to predicting diplomat parking ticket accrual, it also predicts bribery behavior within economic games (Muthukrishna, Francois, Pourahmadi, & Henrich, 2017). We use the 2015 index, the final year of the current wave of the World Values Survey. The Return Wallet data is from a recent study comparing return rates of dropped wallets in several countries around the world (Cohn, Maréchal, Tannenbaum, & Zünd, 2019).

Distance Measures

The CEPII GeoDist dataset (Mayer & Zignago, 2011) contains pairwise geographic distance measures between countries calculated in four different ways. The first (*Geographic Distance Population Center*) is a simple distance calculation based on the latitudes and longitudes of the most important city (in terms of population) in each country. The second (*Geographic Distance Capitals*) takes the same approach, but uses the official capital of each country. The third and fourth measures are weighted to account for the geographic location of the population within each country. This approach, akin to gravitational attraction, calculates the distance between the largest cities in each country, weighted by the share of the population within those cities. *Geographic Distance Gravity Weight 1* uses an arithmetic mean and *Geographic Distance Gravity Weight 2* uses a harmonic mean.

The CEPII Language dataset (Melitz & Toubal, 2014) contains pairwise linguistic distance measures between countries calculated in two different ways. *Linguistic Distance Ethnologue* calculates linguistic distance using the Ethnologue language trees (ethnologue.com).

Linguistic Distance ASJP is based on the results of the Automated Similarity Judgment Program

(ASJP) project (http://asjp.clld.org/), which looks at lexical similarity between 40 words in 256

languages (as defined by the Ethnologue project).

Genetic distance is based on genetic data from Pemberton et al. (2013), matched to

countries by Spolaore and Wacziarg (2016). *Genetic Distance Ethnic Weighting* weights the

ethnic groups within a country by their population size. *Genetic Distance Ethnic Plurality* only

considers genetic distance from the largest ethnic plurality in the country.

Answers as Alleles

The World Values Survey contains various different answer types, including Likert scales,

binary approve/disapprove, categorical responses, etc. Societies have markedly different

answering styles (Chen, Lee, & Stevenson, 1995; Heine, Lehman, Peng, & Greenholtz, 2002).

There are also issues with using Likert scales cross-culturally due to a variety of factors,

including these cross-cultural differences in preferences for Yes or No versus a Likert scale (for

discussion and review, see Hruschka, Munira, Jesmin, Hackman, and Tiokhin, 2018). To reduce

the possibility that differences in answer distributions were due to differences in answering style

rather than the actual belief or behavior of interest, we split and collapsed valenced questions

into positive or negative values, akin to alleles (e.g. "very important" and "rather important"

were combined and "not very important" and "not at all important" were combined; if a mid-

point existed, it was treated as a separate response). This avoided differences in answering styles

masking true differences in beliefs, values, and behaviors and was conservative in collapsing

graded differences and only showing opposite beliefs or behaviors. That is, larger distances may

exist due to the extent of a belief or behavior – for example, both Country A and Country B may

be more liberal, but Country A even more so. Unfortunately, we can't distinguish this from a

tendency for people from Country A to answer more extremely. The distance we show here represents opposite beliefs and behaviors – for example, Country A is more liberal and Country B is more conservative.

Nominal questions such as religious classification ("a religious person", "not a religious person", "a convinced atheist") were treated akin to separate "alleles". We discuss this in more detail below. Moreover, some ordinal options had very few responses for a particular option.

Our decision strategy was as follows:

1. We changed all valence questions to a positive and negative allele. Where a midpoint existed, this was coded as a third allele. Examples:

    a. The question "Indicate how important it is in your life: Family" had answer options "Very important", "Rather important", "Not very important", and "Not at all important". These were condensed to two cultural alleles.

        i. **Allele 1**: "Very important" and "Rather important"

        ii. **Allele 2**: "Not very important" and "Not at all important"

    b. The question "To fully develop your talents, you need to have a job" had answer options "Strongly agree", "Agree", "Neither", "Disagree", and "Strongly disagree". These were condensed to three alleles, one for each valence, and one for the mid-point:

        i. **Allele 1**: "Strongly agree" and "Agree"

        ii. **Allele 2**: "Disagree" and "Strongly disagree"

        iii. **Allele 3**: "Neither"

2. We restricted this breakdown to a maximum of 4 alleles, combining multiple answers based on overall distributions of responses. Examples:

   a. The question "Apart from weddings and funerals, about how often do you attend religious services these days?" had answer options "More than once a week", "Once a week", "Once a month", "Only holy days", "Once a year", "Less often", and "Never" (wording changed slightly between waves). These were condensed to four alleles:

      i. **Allele 1**: "More than once a week", "Once a week"

      ii. **Allele 2**: "Once a month", "Only holy days"

      iii. **Allele 3**: "Once a year", "Less often"

      iv. **Allele 4**: "Never"

A full list of allele categorizations is available in Supplemental Materials *allele-dimension-data.csv.*

Confidence Intervals and Robustness Calculations

We calculated 95% confidence intervals (CI) by bootstrapping with 1000 replications as per Bell et al. (2009). Figure S1 and Figure S2 in the Supplementary Material shows the American scale and Sino scale with 95% CIs.

We investigated the robustness of Cultural F_{ST} in two ways: (a) by randomly resampling a fixed percentage of the question set (increments of 10% from 10% to 90%; new random selection of questions for each sample) and (b) by randomly resampling a fixed percentage of the values (increments of 10% from 10% to 90%; new random selection of values for each sample). In both cases, we compare the calculated CF_{ST} scores to the scores using the full dataset. We did

this 10,000 times for each percentage value for each method of resampling and then compared it

to the real CF_{ST} values, (a) recording the size of the deviation, (b) recording the Pearson

correlation, and (c) recording the Spearman correlation. For the purposes of the American scale,

we did this for all countries relative to the United States. For a comparison, we did the same to

create a Sino scale with all countries relative to China.

Results

American scale and Sino scale

We constructed an American scale by calculating the cultural distance for all countries

relative to the United States, the most overrepresented country in psychological research by a

wide margin (Henrich et al., 2010). As a point of comparison and to reinforce the point that

cultural distances are not unidimensional ranging from WEIRD to non-WEIRD, we also

constructed a Sino scale by calculating the cultural distance for all countries relative to China, a

common cultural comparison in cultural psychology.

Since CF_{ST} is a composite of many questions, we are effectively cutting a line through a

large multidimensional culture-space and only looking at the distance from a particular point (the

United States in the American scale; China in the Sino scale). Thus, Japan and Norway are

similarly culturally distant from the United States (0.115 and 0.124), but are not necessarily

similar to each other, just as Colombia and the United Kingdom are similarly geographically

distant from the United States, but nowhere near each other. Table 1 contains the values of the

American scale and Sino scale of cultural distance, graphed on a map in Figure 3 and number

lines in Figure 4 below. Figures S1 and S2 show the American scale and Sino scale number lines,

respectively, with 95% CIs shaded in blue. To further emphasize that cultural distances are

distances in a large multidimensional culture-space, Figure 5 is a 2D plot of distance from both

the United States and China. The plot reveals that many countries are close to neither the United

States nor China. But here too, bear in mind that countries close to each other on this plot are not

necessarily culturally close to each other; any low dimensional plot will necessarily collapse the

distances in the large multidimensional culture-space.

Recent research (Klein et al., 2018) suggests that researchers may also wish to have a

proxy for a *WEIRD scale*. The American scale may serve as a proxy for a WEIRD scale, given

that American samples dominate psychological research. However, as the American scale

reveals, there is considerable distance between different nations that may all be classified as

WEIRD (see Table 1 and Figures 3 and 4). Moreover, there is a high, but imperfect correlation

between scales constructed with distance from different WEIRD nations (e.g. American scale

correlation with: Canadian scale $r = .94, p < .001$; Australian scale $r = .89, p < .001$; New

Zealand scale $r = .86, p < .001$; British scale $r = .83, p < .001$). Wherever possible, we

encourage researchers to look at the distance between specific samples or even between

participants at a national or regional level. For example, conducting studies to test

generalizability or explain anticipated cultural variation from US studies will be less powerful if

done in Australia compared to Yemen, but generalizing a study conducted in Turkey will be less

powerful if done in Yemen compared to Australia.

To make it easy to quickly calculate the cultural distance between any two nations, we

have provided the *R* code and created an online tool (http://culturaldistance.com). In the future,

this tool will also allow region-by-region comparisons.

Table 1. American Scale and Sino Scale of Cultural Distance from United States

Country	American Cultural Distance	American 95% CI	Sino Cultural Distance	Sino 95% CI
Algeria	0.138	0.132, 0.144	0.221	0.216, 0.228
Andorra	0.115	0.109, 0.122	0.249	0.242, 0.258
Argentina	0.071	0.069, 0.075	0.150	0.146, 0.155
Armenia	0.154	0.149, 0.161	0.177	0.171, 0.183
Australia	0.035	0.033, 0.039	0.131	0.127, 0.135
Azerbaijan	0.175	0.169, 0.181	0.158	0.153, 0.165
Bahrain	0.167	0.161, 0.173	0.195	0.189, 0.201
Belarus	0.071	0.068, 0.075	0.101	0.097, 0.106
Brazil	0.072	0.069, 0.075	0.159	0.156, 0.163
Bulgaria	0.108	0.104, 0.114	0.116	0.111, 0.123
Burkina Faso	0.143	0.139, 0.149	0.153	0.149, 0.157
Canada	0.026	0.025, 0.028	0.135	0.132, 0.140
Chile	0.078	0.075, 0.081	0.156	0.152, 0.161
China	0.150	0.146, 0.155	-	-
Colombia	0.102	0.0987, 0.106	0.182	0.178, 0.186
Cyprus	0.057	0.055, 0.061	0.118	0.114, 0.122
Ecuador	0.109	0.105, 0.114	0.197	0.192, 0.204
Egypt	0.234	0.228, 0.241	0.186	0.183, 0.190
Estonia	0.117	0.112, 0.122	0.097	0.093, 0.102
Ethiopia	0.130	0.126, 0.136	0.153	0.149, 0.158
Finland	0.072	0.069, 0.076	0.176	0.171, 0.185
France	0.079	0.075, 0.085	0.181	0.175, 0.190
Georgia	0.143	0.139, 0.148	0.143	0.140, 0.146
Germany	0.080	0.078, 0.084	0.114	0.111, 0.117
Ghana	0.153	0.149, 0.158	0.172	0.169, 0.175
Great Britain	0.046	0.043, 0.051	0.172	0.166, 0.181
Guatemala	0.134	0.130, 0.140	0.192	0.186, 0.198
Hong Kong	0.090	0.088, 0.095	0.085	0.082, 0.090
Hungary	0.102	0.098, 0.108	0.125	0.121, 0.132
India	0.093	0.091, 0.097	0.106	0.104, 0.110
Indonesia	0.178	0.173, 0.184	0.167	0.163, 0.171
Iran	0.150	0.145, 0.156	0.125	0.122, 0.128
Iraq	0.162	0.158, 0.167	0.193	0.189, 0.197
Italy	0.061	0.059, 0.065	0.163	0.157, 0.169
Japan	0.115	0.112, 0.119	0.118	0.114, 0.122
Jordan	0.195	0.190, 0.200	0.193	0.189, 0.197

Kazakhstan	0.107	0.103, 0.111	0.099	0.095, 0.104
Kuwait	0.122	0.117, 0.127	0.163	0.157, 0.169
Kyrgyzstan	0.132	0.128, 0.137	0.161	0.156, 0.166
Lebanon	0.103	0.099, 0.109	0.175	0.169, 0.182
Libya	0.146	0.142, 0.151	0.198	0.194, 0.202
Malaysia	0.125	0.121, 0.129	0.156	0.153, 0.160
Mali	0.155	0.150, 0.161	0.155	0.151, 0.160
Mexico	0.077	0.074, 0.080	0.138	0.135, 0.141
Moldova	0.100	0.096, 0.105	0.133	0.128, 0.140
Morocco	0.149	0.145, 0.155	0.176	0.172, 0.180
Netherlands	0.079	0.076, 0.083	0.146	0.142, 0.150
New Zealand	0.053	0.050, 0.058	0.162	0.156, 0.168
Nigeria	0.130	0.126, 0.135	0.222	0.217, 0.227
Norway	0.124	0.118, 0.131	0.206	0.199, 0.214
Pakistan	0.178	0.173, 0.185	0.240	0.234, 0.246
Palestine	0.134	0.129, 0.140	0.193	0.187, 0.20
Peru	0.090	0.087, 0.094	0.142	0.139, 0.146
Philippines	0.144	0.139, 0.150	0.229	0.223, 0.236
Poland	0.076	0.074, 0.081	0.147	0.143, 0.151
Qatar	0.176	0.171, 0.183	0.262	0.255, 0.269
Romania	0.103	0.100, 0.108	0.140	0.137, 0.144
Russia	0.085	0.083, 0.088	0.089	0.086, 0.092
Rwanda	0.149	0.145, 0.154	0.143	0.140, 0.146
Serbia and Montenegro[1]	0.079	0.076, 0.084	0.166	0.160, 0.174
Singapore	0.038	0.036, 0.041	0.124	0.120, 0.129
Slovenia	0.077	0.074, 0.081	0.122	0.118, 0.126
South Africa	0.076	0.073, 0.079	0.138	0.135, 0.141
South Korea	0.092	0.089, 0.095	0.073	0.071, 0.077
Spain	0.074	0.071, 0.078	0.137	0.133, 0.142
Sweden	0.115	0.111, 0.121	0.186	0.180, 0.191
Switzerland	0.068	0.064, 0.074	0.179	0.173, 0.187
Taiwan	0.097	0.095, 0.101	0.092	0.089, 0.096
Thailand	0.129	0.125, 0.134	0.104	0.101, 0.107
Trinidad and Tobago	0.088	0.085, 0.093	0.187	0.183, 0.191
Tunisia	0.156	0.151, 0.163	0.179	0.175, 0.185
Turkey	0.120	0.117, 0.126	0.119	0.117, 0.122
Ukraine	0.086	0.083, 0.089	0.117	0.114, 0.123
United States	-	-	0.150	0.146, 0.155
Uruguay	0.084	0.081, 0.088	0.143	0.139, 0.148
Uzbekistan	0.150	0.146, 0.155	0.150	0.146, 0.155
Vietnam	0.182	0.177, 0.188	0.057	0.055, 0.061

Yemen	0.200	0.193, 0.209	0.248	0.241, 0.256
Zambia	0.083	0.081, 0.088	0.162	0.158, 0.167
Zimbabwe	0.110	0.106, 0.115	0.220	0.216, 0.226

[1] Serbia and Montenegro separated in 2006, however, they remain combined in the World Values Survey data.
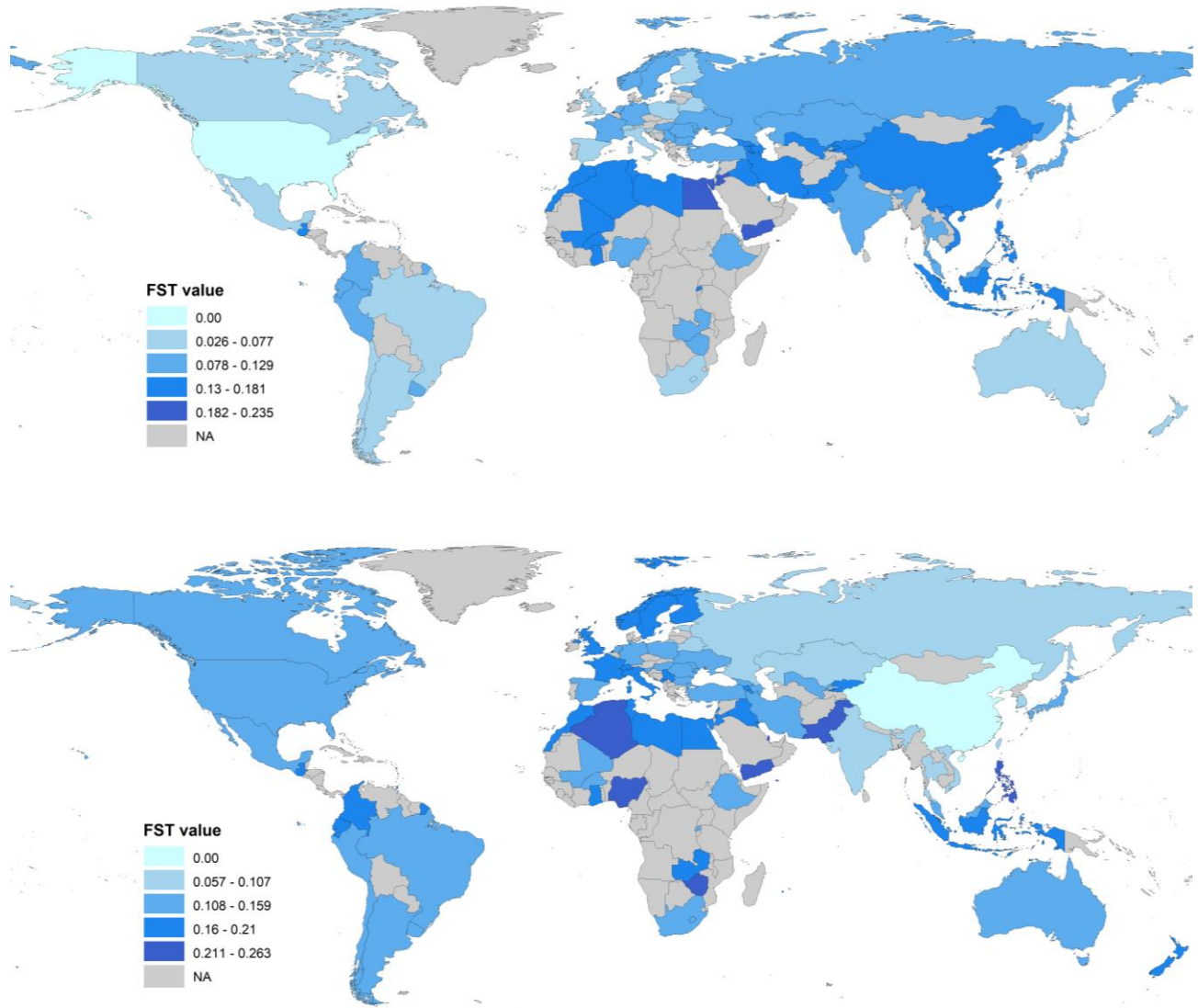

Figure 3. Top: American Scale of Cultural Distance from the United States visualized on a world map. Bottom: Sino Scale of Cultural Distance from China visualized on a world map.
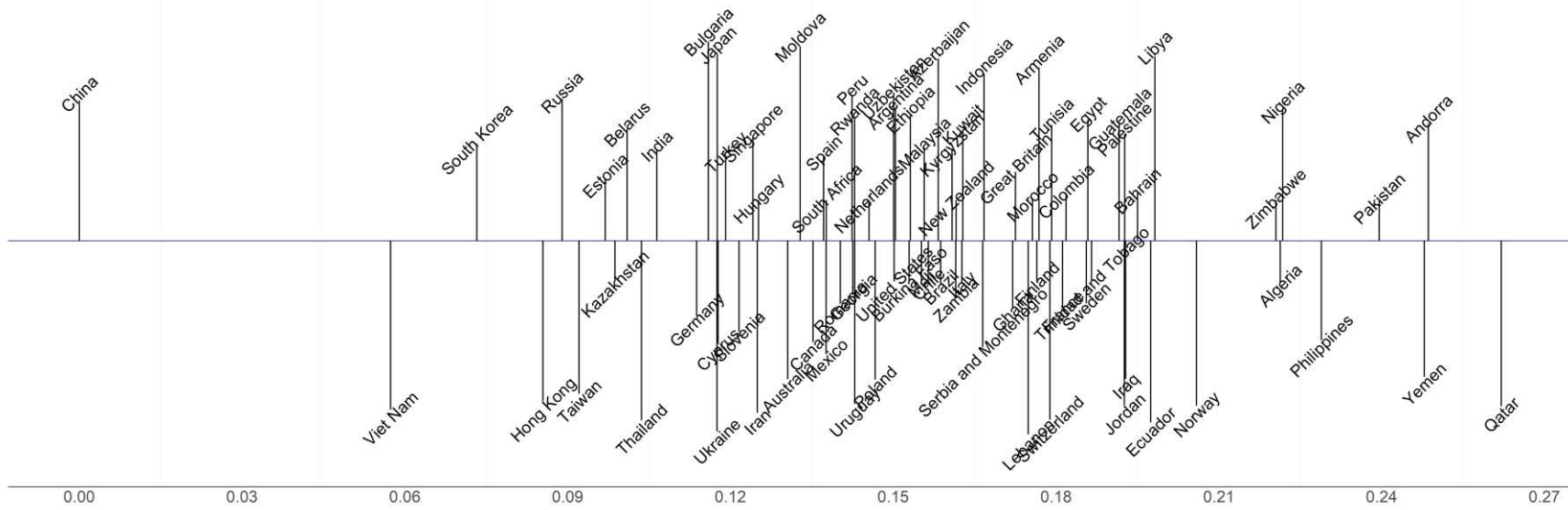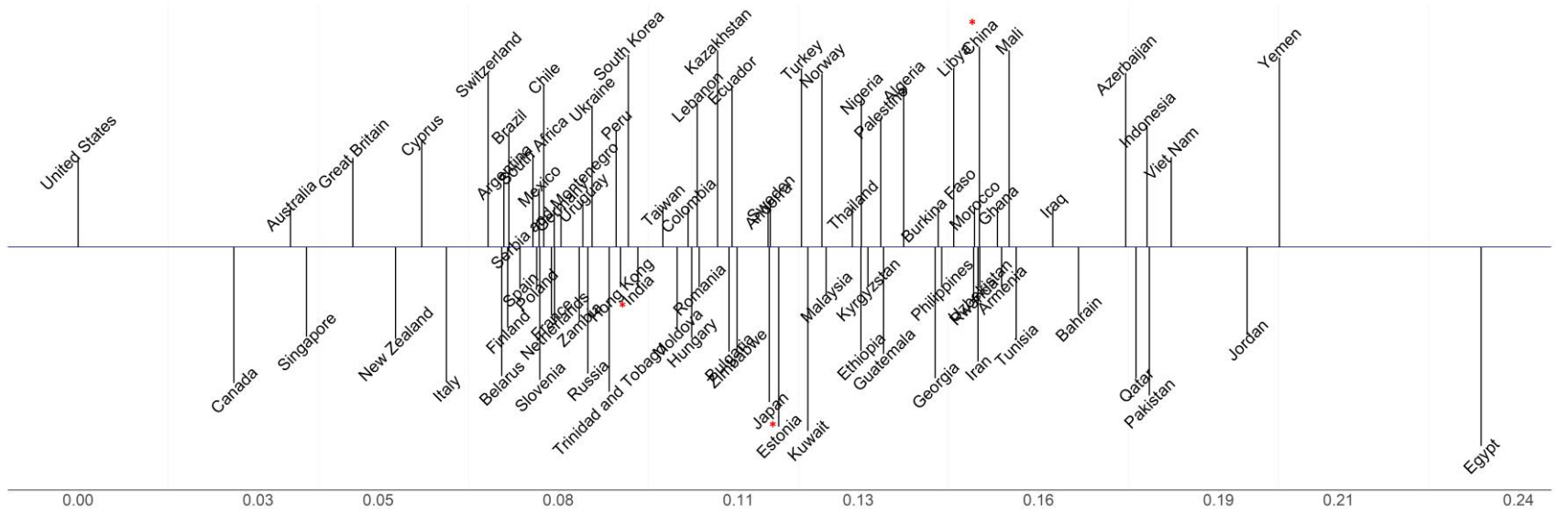
Figure 4. Top: American Scale of Cultural Distance from the United States. Note that the most commonly studied non-Western nations - Japan, Hong Kong, and China (marked with asterisks) - are by no means the extreme on the American scale. Bottom: Sino Scale of Cultural Distance from China.
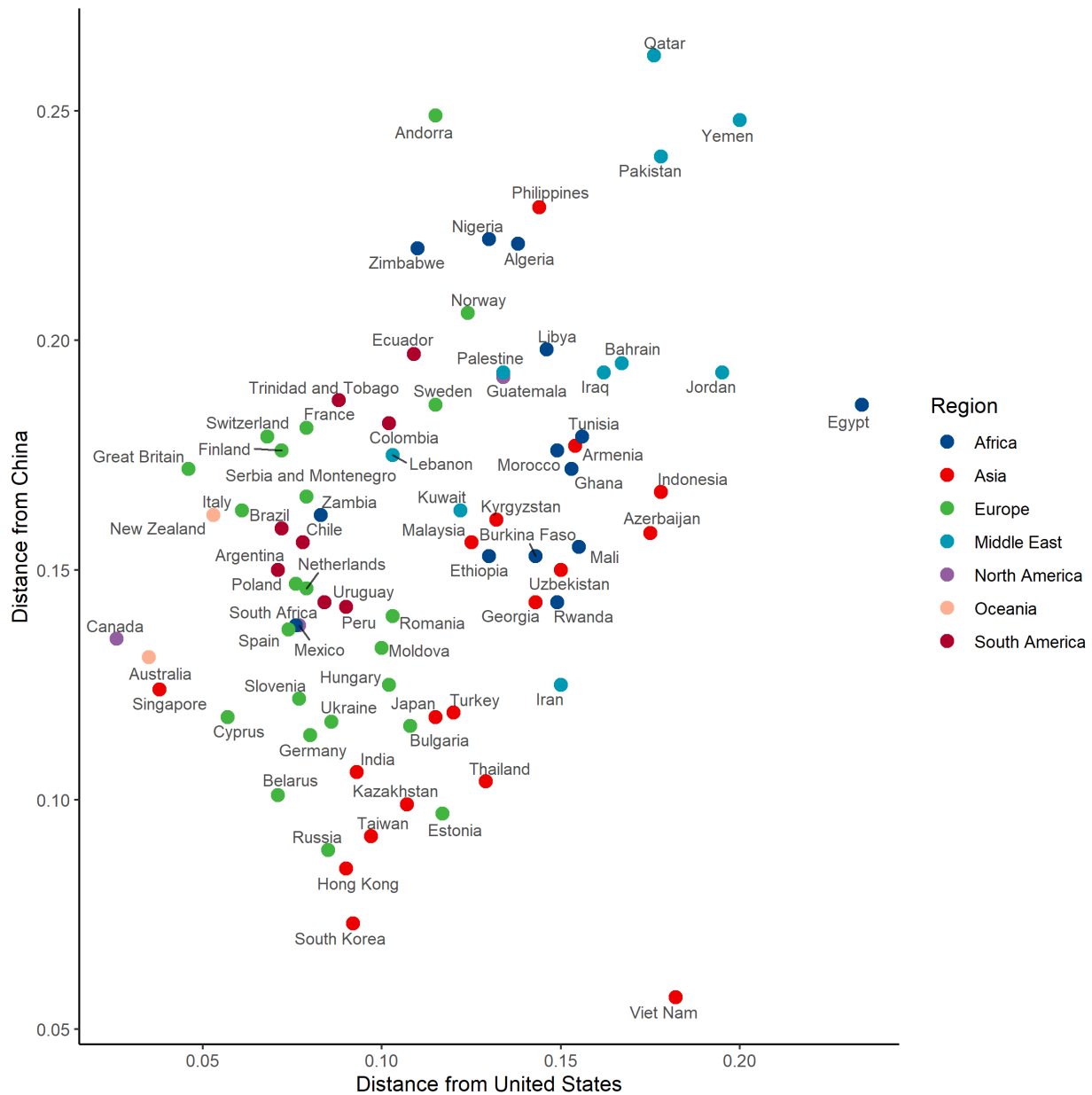
Figure 5. Plot of CF_{ST} from United States and China. Countries closer to the top of the graph are culturally further away from China; being close to the x-axis suggests closeness to China. Countries further right on the graph are culturally further away from the United States; being close to the y-axis suggests closeness to the United States. This plot shows the multidimensional nature of cultural distance – many countries are close to neither the United States nor China (and are not necessarily culturally close to each other). For example, Sweden and Colombia are similarly distant from both the United States (0.115 and 0.102) and China (0.186 and 0.182), but they are not culturally close to each other (indeed, they are very different CF_{ST} = 0.261).

Within-nation regional variation

Many readers will be familiar with the cultural differences within the United States – the honor culture of the South (Nisbett & Cohen, 1996), the corporate and educational culture of New England, the liberal culture of the West Coast, and so on (Woodard, 2011). However, CF_{ST} analyses comparing the WVS regions (with at least 100 surveyed individuals) within the United States, China, India, and the European Union (EU) reveals the relative homogeneity of the United States compared to these other large populations.

India, "the country of a hundred nations and a hundred tongues, of a thousand religions and two million gods" (Mark Twain, "Following the Equator") has the largest mean regional diversity of these populations ($CF_{ST}= 0.11$); followed by the EU, a long-running project attempting to create a political and economic union ($CF_{ST}= 0.09$); followed by China, despite being 90% Han (CFST = 0.05). The United States has the least regional diversity with $CF_{ST} = 0.01$.

We illustrate these differences in Figure 6 below, conducting a non-metric multidimensional scaling (NMDS) with the *R MASS* package on a pairwise CF_{ST} matrix including the WVS regions within the United States, China, India, and the EU with at least 100 surveyed individuals.

These within-population analyses reveal the importance of remembering that societies are not homogenous, but rather are multivariate distributions of many traits along many dimensions with structure within structure. There are likely to be cultural differences not only between regions within a country, but also ethnicities, religions, socioeconomic class, and other groupings. These are all avenues for future research.
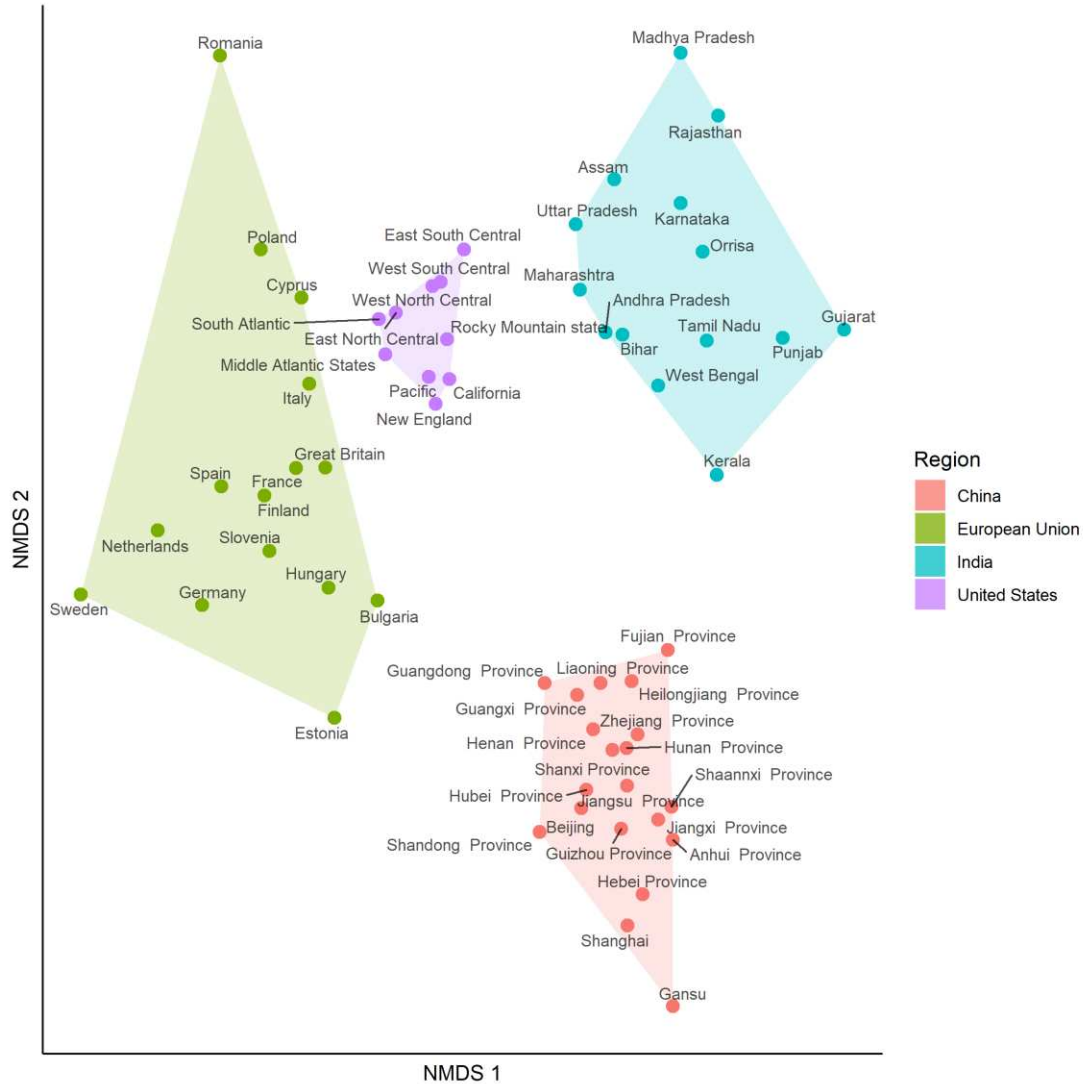
Figure 6. Two-dimensional non-metric Multidimensional Scaling (NDMS) plot of the pairwise CF_{ST} matrix of regions with at least 100 survery individuals in the four largest populations: United States, China, India, and the countries of the European Union.

Robustness tests

We investigated the robustness of CF_{ST} by randomly resampling a fixed percentage of the question set and comparing these values to the full question set. We did this 10,000 times for each percentage value and then (a) counted the number of times a value fell outside the confidence interval calculated on all questions and (b) recorded the size of the deviation from the

CF_{ST} value calculated on all questions for two different measures of robustness. For the purposes of the American scale we did this for all countries relative to the United States. For the Sino scale, we did the same with all countries relative to China. We plot these in Figures 7 and 8 below.
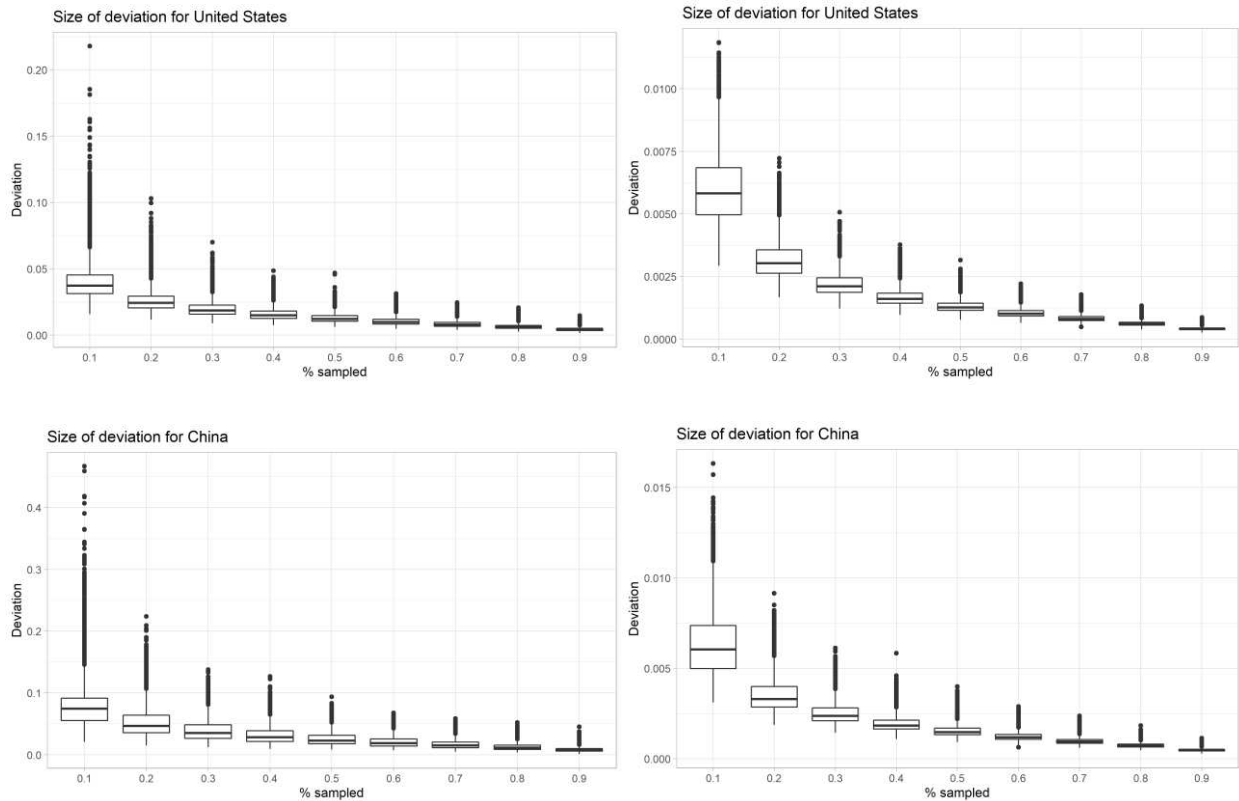

Figure 7. Robustness of American and Sino scale showing the size of the deviations in the 10,000 resampled values for each percentage sample. Left: Sampling by percentage of questions. Right: Sampling by percentage of values. Removing entire questions has a larger effect on the final values than removing values at random, but the effect of loss of data is still small for up to 50% of questions removed.

Figure 8. Robustness of American and Sino scale showing correlations in the 10,000 resampled values for each percentage sample. Left: Sampling by percentage of questions. Right: Sampling by percentage of values. Removing entire questions has a larger effect on the final values than removing values at random, but the effect of loss of data is still small for up to 50% of questions removed.
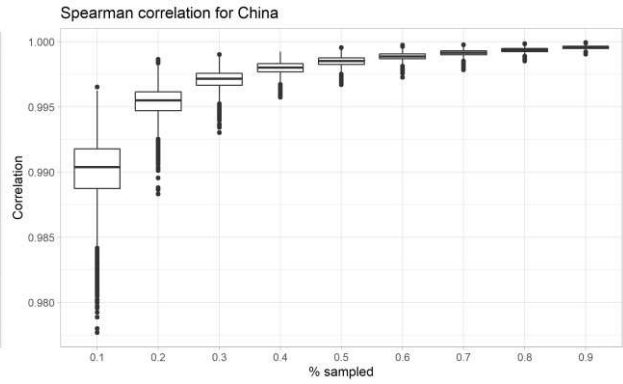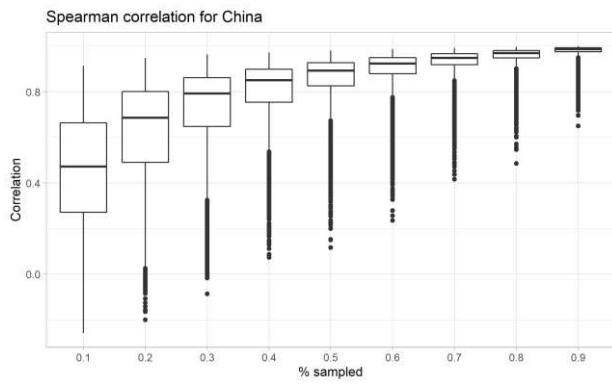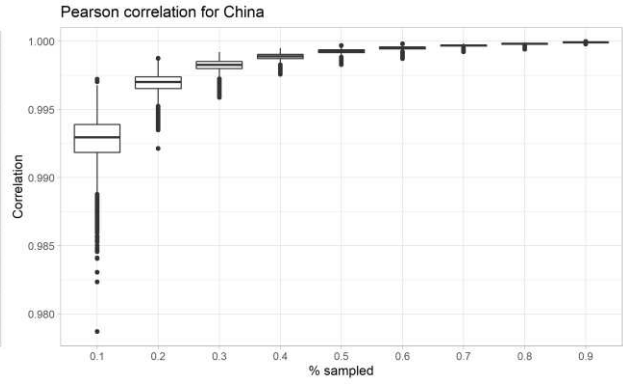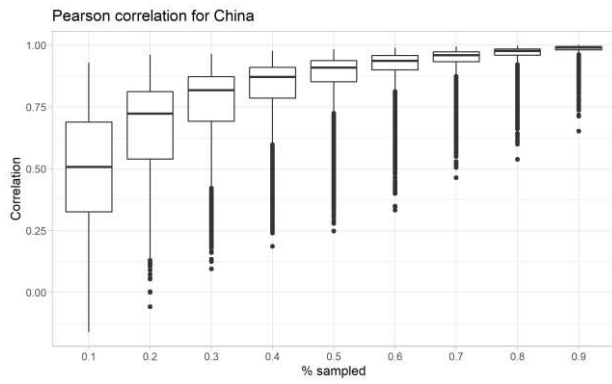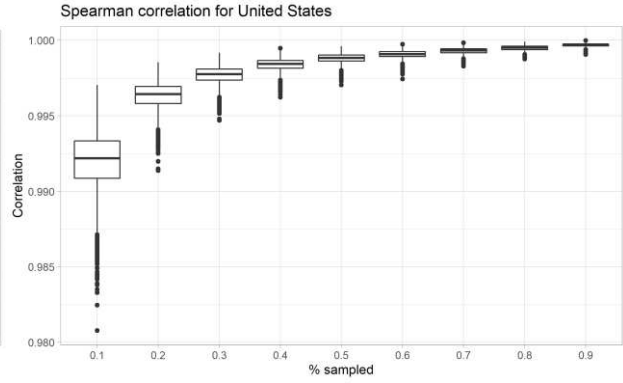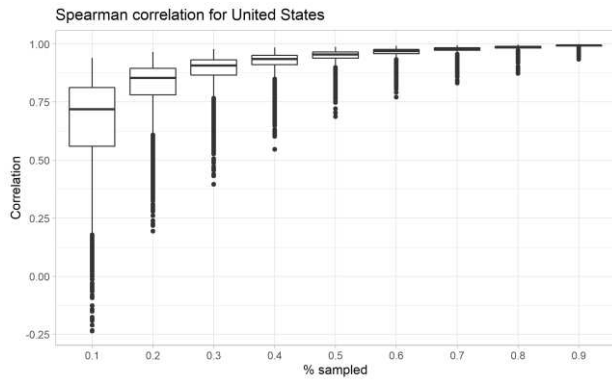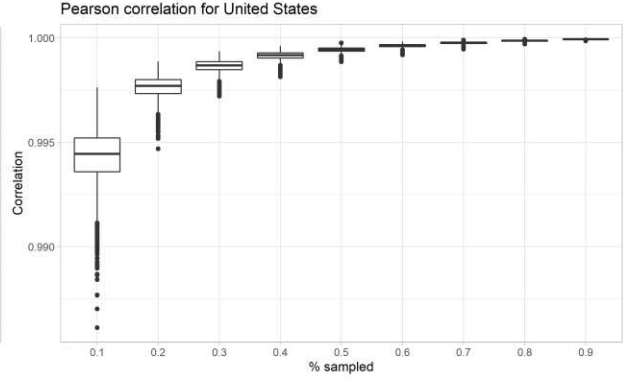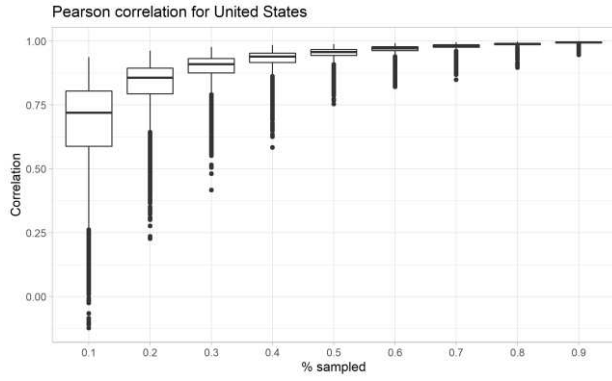
These analyses show that CF_{ST} is highly robust. We can sample up to only 50% of questions and still get only small deviations (0.013 and 0.026 from US and China respectively) and values that remain highly correlated with the original scale by both Pearson's correlation ($r = 0.95$ and $r = 0.87$ from US and China respectively) and Spearman's correlation ($r = 0.95$ and $r = 0.85$ from US and China respectively). When sampling by values instead of entire questions, deviations are smaller still. This robustness, even to losing entire questions, is not so surprising when you consider that culture tends to cluster (Harton & Bullock, 2007), due to cultural transmission mechanisms such as common sources of information and the conformist bias in cultural learning (Chudek, Muthukrishna, & Henrich, 2015; Henrich, 2016; Muthukrishna, Morgan, & Henrich, 2016). Thus, CF_{ST} will be robust even if we don't ask every conceivable question, as long as we have a variety of questions that capture a variety of cultural traits (the correlation between cultural traits is not perfect). To this end, the WVS is an ideal survey for creating scales of cultural distance.

Comparison to other measured psychological and cultural differences

Here we compare the American scale and Sino scale to other commonly used measures of psychological and cultural differences. For the difference measures, we report the correlation with the raw value and also subtract each country's score from the score for the United States and China respectively (labelled "Relative"). For the distance scores, we compare the distance from the United States and China, respectively. Table 2 lists these correlations. We plot these in Figure S4.

These correlations reveal that the American scale predicts many cross-cultural differences in psychology, more so than the Sino scale. Remarkably, many of these correlations are large and significant even when correlating the raw values of the various measures. This suggests that WEIRD societies, as typified by the United States, are truly odd outliers in human psychology (Schulz et al., 2019). Another non-exclusive possibility is that researchers from WEIRD countries have focused their data gathering efforts on psychological dimensions that they find odd in comparison to Western psychology. Long term orientation, a new addition to Hofstede's cultural dimensions, is a nice illustration of a trait where China, rather than the US, is the outlier. If psychology were dominated by China, Table 2 would have perhaps been dominated by dimensions in which Chinese researchers found different in other parts of the world and the Sino scale, rather than the American scale, would consequently be more predictive of differences.

In contrast to these psychological differences, it is reassuring to see that the proxies for cultural distance – geographic, linguistic, and genetic – though weakly correlated – are predictive of cultural distance from both the United States and China and not just the United States. Together these correlations suggest that (1) we are measuring an overall cultural distance—a strength of the CF_{ST} approach; (2) that an American scale is predictive of various psychological variables; and (3) the United States is unique either as a psychological outlier or due to what US and other WEIRD researchers have chosen to study.

Table 2. Correlation between American scale (CF$_{ST}$ distance from the United States), Sino scale (CF$_{ST}$ distance from China) and other commonly used psychological and cultural difference and distance measures.

Correlated Measure	American Scale r	American Scale [95% CI]	Sino Scale r	Sino Scale [95% CI]	df
Hofstede's Cultural Dimensions					
Individualism[1]	-0.51***	[-0.68,-0.30]	-0.02	[-0.27, 0.24]	57
Individualism Relative[1]	0.51***	[0.30, 0.68]	0.06	[-0.20, 0.31]	57
Power Distance[1]	0.42***	[0.19, 0.61]	0.06	[-0.20, 0.31]	57
Power Distance Relative[1]	0.40**	[0.16, 0.60]	0.00	[-0.25, 0.26]	57
Masculinity[1]	-0.06	[-0.32, 0.19]	0.07	[-0.19, 0.32]	57
Masculinity Relative[1]	0.06	[-0.20, 0.31]	0.20	[-0.06, 0.44]	57
Uncertainty Avoidance[1]	0.00	[-0.25, 0.26]	-0.03	[-0.29, 0.22]	57
Uncertainty Avoidance Relative[1]	-0.06	[-0.31, 0.20]	-0.29*	[-0.50,-0.03]	57
Long Term Orientation[1]	-0.23+	[-0.46, 0.04]	-0.55***	[-0.71,-0.34]	53
Long Term Orientation Relative[1]	-0.06	[-0.32, 0.21]	0.53***	[0.31, 0.70]	53
Indulgence[1]	-0.44***	[-0.63,-0.20]	0.20	[-0.07, 0.44]	53
Indulgence Relative[1]	0.50***	[0.27, 0.67]	0.33*	[0.07, 0.55]	53
Tightness-Looseness					
Tightness Gelfand[2]	0.41*	[0.02, 0.70]	0.11	[-0.30, 0.48]	23
Tightness Gelfand Relative[2]	0.62**	[0.29, 0.81]	0.21	[-0.20, 0.56]	23
Looseness Composite[3]	-0.75***	[-0.86,-0.57]	-0.12	[-0.41, 0.19]	38/39
Looseness Composite Relative[3]	0.54***	[0.28, 0.73]	0.18	[-0.13, 0.47]	38/39
Looseness Domain Specific[3]	-0.71***	[-0.83,-0.52]	-0.39*	[-0.62,-0.10]	40
Looseness Domain Specific Relative[3]	0.73***	[0.55, 0.85]	0.06	[-0.25, 0.36]	40
Looseness Domain General[3]	-0.81***	[-0.90,-0.66]	-0.30+	[-0.56, 0.01]	38/39
Looseness Domain General Relative[3]	0.69***	[0.48, 0.82]	-0.11	[-0.40, 0.21]	38/39

Schwartz's Values	Harmony Value[4]	-0.25[+]	[-0.49, 0.03]	-0.06	[-0.32, 0.22]	50
	Harmony Cultural Relative[4]	-0.25[+]	[-0.49, 0.03]	-0.09	[-0.35, 0.19]	50
	Mastery Value[4]	-0.06	[-0.32, 0.22]	-0.21	[-0.46, 0.07]	50
	Mastery Value Relative[4]	0.25[+]	[-0.03, 0.49]	0.21	[-0.07, 0.46]	50
	Embeddedness Value[4]	0.66[***]	[0.47, 0.79]	0.21	[-0.06, 0.46]	50
	Embeddedness Value Relative[4]	0.56[***]	[0.33, 0.72]	0.58[***]	[0.37, 0.74]	50
	Hierarchy Value[4]	0.27[+]	[-0.00, 0.50]	-0.22	[-0.46, 0.06]	50
	Hierarchy Value Relative[4]	-0.03	[-0.30, 0.25]	0.22	[-0.06, 0.46]	50
	Egalitarianism Value[4]	-0.40[**]	[-0.61,-0.15]	0.27[*]	[0.00, 0.51]	50
	Egalitarianism Value Relative[4]	-0.10	[-0.36, 0.18]	0.27[+]	[-0.00, 0.50]	50
	Affective Autonomy Value[4]	-0.57[***]	[-0.73,-0.35]	-0.16	[-0.42, 0.11]	50
	Affective Autonomy Value Relative[4]	0.55[***]	[0.32, 0.71]	0.39[**]	[0.13, 0.60]	50
	Intellectual Autonomy Value[4]	-0.49[***]	[-0.67,-0.25]	-0.15	[-0.41, 0.13]	50
	Intellectual Autonomy Value Relative[4]	-0.16	[-0.42, 0.11]	0.34[**]	[0.07, 0.56]	50
Five Factor Model of Personality	Openness[5]	-0.29[+]	[-0.57, 0.05]	0.16	[-0.18, 0.47]	33
	Openness Relative[5]	-0.15	[-0.46, 0.19]	-0.01	[-0.34, 0.33]	33
	Conscientiousness[5]	-0.09	[-0.41, 0.25]	-0.04	[-0.37, 0.29]	33
	Conscientiousness Relative[5]	0.18	[-0.16, 0.49]	0.20	[-0.15, 0.50]	33
	Extraversion[5]	-0.53[**]	[-0.73,-0.23]	-0.12	[-0.44, 0.22]	33
	Extraversion Relative[5]	0.48[**]	[0.18, 0.70]	-0.00	[-0.34, 0.33]	33
	Agreeableness[5]	-0.33[+]	[-0.59, 0.01]	-0.17	[-0.48, 0.17]	33
	Agreeableness Relative[5]	-0.00	[-0.33, 0.33]	0.15	[-0.19, 0.46]	33
	Neuroticism[5]	-0.09	[-0.42, 0.25]	0.09	[-0.25, 0.41]	33

Neuroticism Relative[5]	-0.06	[-0.39, 0.28]	0.09	[-0.25, 0.41]	33
Personality Std. Dev.[5]	-0.40*	[-0.64,-0.07]	-0.19	[-0.49, 0.16]	33
Personality Std. Dev. Relative[5]	0.53**	[0.23, 0.73]	-0.01	[-0.34, 0.33]	33

Other Psychological and Behavioral Measures

Blood Donations[6]	-0.50***	[-0.66,-0.29]	-0.34*	[-0.54,-0.11]	63
Blood Donations Relative[6]	0.51***	[0.31, 0.67]	-0.29*	[-0.50,-0.05]	63
Diplomat Parking Tickets[7]	0.40***	[0.19, 0.58]	0.14	[-0.10, 0.37]	67
Diplomat Parking Tickets Relative[7]	Data not available		0.16	[-0.08, 0.39]	67
Corruption CPI[8]	-0.50***	[-0.65,-0.31]	-0.15	[-0.36, 0.08]	74
Corruption CPI Relative[8]	0.47***	[0.28,-0.63]	-0.03	[-0.25, 0.20]	74
Return Wallet without Money[9]	-0.53**	[-0.75,-0.21]	0.32+	[-0.05, 0.61]	31
Return Wallet without Money Relative[9]	0.45*	[0.10, 0.70]	0.32+	[-0.05, 0.61]	31
Return Wallet with Money[9]	-0.49**	[-0.72,-0.16]	0.23	[-0.15, 0.54]	31
Return Wallet with Money Relative[9]	0.51**	[0.18, 0.73]	0.23	[-0.14, 0.54]	31

Distance Measures

Kogut-Singh Cultural Distance[10]	0.41**	[0.17, 0.60]	0.01	[-0.24, 0.27]	57
Kogut-Singh Cultural Distance All[10]	0.43**	[0.18, 0.62]	0.37**	[0.11, 0.58]	57
Geographic Distance Population Center[11]	0.21+	[-0.01, 0.42]	0.25*	[0.02, 0.45]	72
Geographic Distance Capitals[11]	0.23*	[0.00, 0.44]	0.25*	[0.03, 0.46]	72
Geographic Distance Gravity Weight 1[11]	0.29*	[0.06, 0.48]	0.26*	[0.04, 0.46]	72
Geographic Distance Gravity Weight 2[11]	0.29*	[0.07, 0.49]	0.27*	[0.05, 0.47]	72
Linguistic Distance Ethnologue[12]	0.14	[-0.17, 0.43]	Not enough data		38
Linguistic Distance ASJP[12]	-0.17	[-0.39, 0.08]	0.14	[-0.10, 0.38]	65/63
Genetic Distance Ethnic Weighting[13]	0.21+	[-0.02, 0.42]	0.37**	[0.16, 0.55]	72
Genetic Distance Ethnic Plurality[13]	0.17	[-0.07, 0.38]	0.38***	[0.17, 0.56]	72

*** p < .001   ** p < .01   * p < .05   + p < .10

[1] Scraped from geert-hofstede.com. Higher scores indicate greater values on the raw scale. Relative values are absolute values relative to the comparison country (United States or China).

[2] Gelfand et al. (2011). Higher scores indicate greater tightness. The mean of East and West Germany was used for Germany. Relative values are absolute values relative to the comparison country (United States or China).

[3] Uz (2015). This is a measure of looseness that uses variance in World Values Survey responses instead of the Gelfand et al. (2011) scale. Higher scores indicate greater looseness. Relative values are absolute values relative to the comparison country (United States or China). The domain general and composite values did not exist for China. For the relative measure we use the domain specific value as a proxy.

[4] Schwartz's Culture Value Orientation Scores (Schwartz, 2006). The mean of East and West Germany was used for Germany. The mean of French and German Switzerland was used for Switzerland. Relative values are absolute values relative to the comparison country (United States or China).

[5] The personality factor data for each country was taken from Table 2 in McCrae et al. (2005). The mean of French and German Switzerland was used for Switzerland. Relative values are absolute values relative to the comparison country (United States or China).

[6] Blood Donations data per 1000 persons data was collated from the WHO Global Status on Blood Safety and Availability by (Schulz et al., 2019).

[7] Unpaid parking ticket accrued by diplomats in New York City from Fisman and Miguel (2007).

[8] Corruption perceptions index is a measure of the descriptive corruption norm from Transparency International's 2015 report.

[9] Percentage of dropped wallets with money returned taken from Figure 1 in Cohn et al. (2019).

[10] Cultural distance was calculated as per Kogut and Singh (1988) on the original four Hofstede dimensions (power distance, individualism, masculinity, and uncertainty avoidance; labelled Original) and on all six dimensions, (labelled All).

[11] http://www.cepii.fr GeoDist database (Mayer & Zignago, 2011). Higher scores indicate a larger distance.

[12] http://www.cepii.fr Language database (Melitz & Toubal, 2014). Higher scores indicate greater difference in language.

[13] Based on genetic data from Pemberton et al. (2013) matched to country by Spolaore and Wacziarg (2016). Higher scores indicate a larger genetic distance.

Note: In an earlier version of this paper, we also showed that both the American scale and Sino scale predict the importance of obedience and the American scale and not the Sino scale predict the value of creativity. These items were identified by Schulz et al. (2019) as part of a WEIRD package (specifically, low obedience and high creativity). However, since these are derived from WVS questions, we've removed them from this final version.

Hofstede's individualism dimension is perhaps the dimension most often used and cited across the social sciences. It has the highest correlation with the American scale of cultural distance, with more collectivist societies further away from the United States. The next most commonly used and cited cultural dimension is probably power distance, which has the second highest correlation with the American scale. Societies with a larger power distance are more culturally distant to the United States. The only other measure with a reasonably large correlation with the American scale is indulgence, where societies that have more emphasis on individual freedom are more culturally similar to the United States. These correlations match US stereotypes that emphasize equality and individual freedom. The American scale has a smaller correlation with long term orientation, where societies with more focus on the future are more culturally similar to the United States. The Sino scale only correlates with long term orientation and perhaps indulgence. These results may suggest that these dimensions largely emphasize cultural differences in psychology that look remarkable from a WEIRD standpoint, but perhaps not from the standpoints of other societies.

The tightness-looseness scale developed by Gelfand has a moderate correlation with the American scale—tighter societies are more culturally distant from the United States, but there are several outliers. In contrast, the Uz measure is strongly correlated with the American scale of

cultural distance, with tighter societies more culturally distant from the United States. However, with the possible exception of the domain general Uz measure, tightness-looseness does not reliably correlate with the Sino scale. This suggests that internal cultural variation or tolerance for deviation may be distinct features of the United States, remarkable from a WEIRD standpoint, but not representative of the world.

Schwartz's values of embeddedness and autonomy correlate with the American scale, as may harmony and hierarchy. These correlations mirror the correlation with Hofstede's individualism and indulgence dimensions. The correlations with the Sino scale are more unreliable, though there may be a relationship with embeddedness, egalitarianism, and autonomy, suggesting that China may be an extreme on these dimensions.

Extraversion, Agreeableness, and perhaps Openness correlated with the American scale, again suggesting the United States as an outlier on these dimensions. None of the personality factors reliably correlate with the Sino scale. McCrae et al. (2005) noted that Western nations tend to have larger variance across personality traits, which is borne out by the correlation with the American scale, but not the Sino scale. This may fit with greater latitude for self-expression in WEIRD societies.

WEIRD nations are outliers on creativity, altruism, obedience, and corruption (Schulz et al., 2019). In contrast, the Sino scale correlates with obedience and blood donations, but not the other psychological outcomes.

All distance measures show weak to moderate correlations with the American scale. With the exception of the traditional Kogut-Singh scale and the language distance scale, most also correlate with the Sino scale. The traditional Kogut-Singh scale derived from the original four Hofstede dimensions does not correlate with the Sino scale, but the addition of the two variables

that correlate with the Sino scale increases the correlation. These correlations are reassuring in showing that CF_{ST} approach to measuring cultural distance correlates in the right direction with less high-resolution measures of cultural distance. The relationship between linguistic distance and the American scale was inconsistent, though the plots make clear that there is a huge amount of variance. Many countries with linguistic similarity are quite culturally distant and vice versa. Both genetic measures suggest a fairly modest correlation with our American scale. Such a low value is problematic given the common usage of genetic distance as proxy for cultural distance (Gorodnichenko & Roland, 2017), but is consistent with other researchers (Giuliano, Spilimbergo, & Tonon, 2013), who argue that genetic distance captures geography, but not culture.

Discussion

The psychological sciences face multiple crises, one of which is its overreliance on samples from Western Educated Industrialized Rich Democratic (WEIRD) nations and on samples from the United States in particular (Henrich et al., 2010). A more general theory of human behavior requires a theoretical and empirical understanding of humans across the globe and across the lifespan. We present a cultural distance metric based on the fixation index (F_{ST}) technique from population genetics, applied to the World Values Survey (WVS), a large survey of cultural values. Cultural F_{ST} (CF_{ST}) is a theoretically defensible and robust method of measuring cultural distance, grounded in evolutionary theory. It considers differences between distributions of cultural traits rather than point estimates or arbitrary dimensions. This approach has proved useful in answering questions in anthropology and economics (Bell et al., 2009;

Desmet et al., 2017; Spolaore & Wacziarg, 2016). We anticipate that it will be equally useful to psychologists in addressing the WEIRD people problem.

To this end, we use CF_{ST} to develop an American scale of cultural distance from the United States and a Sino scale of cultural distance from China. CF_{ST} values can range from 0 to 1, yet all values on both scales are less than 0.3. This is consistent with past research (e.g. Saucier et al., 2015) showing that we have more in common across cultures than we have differences. Yet, just as only 4% of our genes separate us from chimpanzees, those differences can be important and predictive.

The American scale correlates with many documented cross-cultural psychological differences, but the Sino scale is far less predictive. Remarkably, the American scale correlates even with the raw scores of these various measures. Together these results suggest that WEIRD nations are truly psychological outliers in some objective sense, as has been recently argued (Schulz et al., 2019). Alternatively, though not mutually exclusively, it could be that these psychological measures have been studied because they are remarkable to researchers from WEIRD countries or remarkable because of how they differ from other nations. That is, if psychology was dominated by Chinese psychologists, we would see a different set of psychological outcomes covered in textbooks, and these psychologies would correlate with a Sino scale. Resolving which of these explanations is correct will require greater diversity in both researchers and samples.

Reassuringly, both the American scale and the Sino scale correlate with proxies of cultural distance, suggesting that the scales are capturing some true cultural distance. In contrast, the original Kogut-Singh distance is only predictive of the American scale. The addition of new dimensions with strong correlations to the Sino scale improve the correlation with both the

WEIRD and Sino scale. As we argue, in capturing only mean differences and ignoring

differences in distributions, this alternative approach is systematically misleading.

CF_{ST} (which has been made available through R code and at http://culturaldistance.com)

allows researchers to measure cultural distance between any two countries. The technique may

also be used to explore cultural differences between regions within countries, between social

classes, between age groups, or any other grouping. We use this flexibility to compare the

cultural differences between regions of the four largest populations—China, India, United States

and European Union. These analyses reveal that the cultural differences between regions of the

over-scrutinized United States are considerably smaller than the European Union, China, or

India.

Researchers may disagree with our various decisions in constructing these scales. For

example, these scales are conservative in removing much of the variance in the degree to which

societies agree or disagree on a cultural trait. We removed these to reduce concerns around

response biases (see *Answers as Alleles*), but other researchers may wish to include the full

extent of this variance. Similarly, other researchers may wish to use a different set of questions

they believe are more defensible as culturally transmissible (see *included-variables.csv*) or even

argue that a different statistic is more appropriate (see Section 2 of Supplementary Materials).

And indeed the World Values Survey itself may have biases in the questions chosen. We hope

that the code we provide and the transparency of our decisions aid in extending this research

beyond these limitations.

We hope that this technique and tool may guide researchers in selecting sites and samples

that are sufficiently culturally different to test the generalizability of their hypotheses. For

example, the Far East has always held a certain exoticism for those from the West, which may

have driven a generation of cultural psychologists to document the many ways in which East

Asian societies differ from the West. However, as Figure 4 illustrates, the most extensively

researched East Asian nations aren't anywhere near the extreme on the American scale and some

are barely halfway. Moreover, as illustrated in Figure 6, there is considerable diversity within

China, let alone between China, Japan, and Hong Kong. This diversity has been exploited by

some researchers; for example, showing the role of agriculture on individualism and collectivism

within China (Talhelm et al., 2014). But we know far less about psychological differences within

countries beyond the United States, where we know state-by-state differences in psychological

traits such as tightness-looseness (Harrington & Gelfand, 2014). We hope that researchers find

the CF_{ST} technique, toolkit, and American scale useful for not only generalizing their findings,

but also for developing theories to explain cross-cultural differences between and within nations.

With such theories, it may also make more sense to use dimensions of cultural distances rather

than an aggregate scale, but we emphasize the need for this investigation to be theoretically

driven (Muthukrishna & Henrich, 2019). In the Supplementary Material, we offer some

suggestions for developing these dimensions.

Relatively little attention has been paid to the Middle East and Africa by both the World

Values Survey (see gray regions in Figure 3) and the psychological sciences more generally.

However, given the relative cultural distance from the United States and Africa's large genetic

(Ramachandran et al., 2005), linguistic (Atkinson, 2011), and likely cultural variation, we have

every reason to suspect the American scale will continue to stretch as we map out these

psychological *terrae incognitae*. These regions, as well as other underrepresented regions, such

as the South Pacific, may in fact hold a treasure trove of findings for the next wave of cultural

psychologists. And as our results illustrate, this may not only shape the breadth of existing

psychological outcomes, but also lead to questions we haven't even thought to ask, new

psychologies, and new ways of organizing psychology. Thus, what we know so far may represent

the tip of the iceberg of a more fully-fledged picture of the human psyche.

Author Contributions

M. Muthukrishna developed the study concept with guidance from A.V. Bell and J. Henrich. A.V. Bell provided advice and analysis code for fixation index. M. Muthukrishna, C. M. Curtin, A. Gedranovich, J. McInerney, and B. Thue wrote code for further analysis and visualization. M. Muthukrishna drafted the paper. All authors provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgments

References

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science, 332*(6027), 346-349.

Bell, A. V., Richerson, P. J., & McElreath, R. (2009). Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proceedings of the National Academy of Sciences, 106*(42), 17671-17674.

Boyd, R. (2017). *A different kind of animal: how culture transformed our species: how culture transformed our species*: Princeton University Press.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*: Princeton university press.

Chen, C., Lee, S.-y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science, 6*(3), 170-175.

Chudek, M., Muthukrishna, M., & Henrich, J. (2015). Cultural Evolution. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (2nd ed., Vol. 2). Hoboken, NJ: John Wiley and Sons.

Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science, 365*(6448), 70-73.

Demes, K. A., & Geeraert, N. (2014). Measures Matter:Scales for Adaptation, Cultural Distance, and Acculturation Orientation Revisited. *Journal of Cross-Cultural Psychology, 45*(1), 91-109.

Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2017). Culture, ethnicity, and diversity. *American Economic Review, 107*(9), 2479-2513.

Fisman, R., & Miguel, E. (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political economy, 115*(6), 1020-1048.

Gelfand, M. (2019). *Rule Makers, Rule Breakers: Tight and Loose Cultures and the Secret Signals That Direct Our Lives*: Scribner.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., . . . Arnadottir, J. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332*(6033), 1100-1104.

Giuliano, P., Spilimbergo, A., & Tonon, G. (2013). Genetic distance, transportation costs, and trade. *Journal of Economic Geography, 14*(1), 179-198.

Gorodnichenko, Y., & Roland, G. (2017). Culture, institutions, and the wealth of nations. *Review of Economics and Statistics, 99*(3), 402-416.

Harrington, J. R., & Gelfand, M. J. (2014). Tightness–looseness across the 50 united states. *Proceedings of the National Academy of Sciences*, 201317937.

Harton, H. C., & Bullock, M. (2007). Dynamic social impact: A theory of the origins and evolution of culture. *Social and Personality Psychology Compass, 1*(1), 521-540.

Heine, S. J. (2015). *Cultural Psychology* (3rd ed.): WW Norton & Company.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology, 82*(6), 903.

Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61-83.

Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). New York, NY: Sage Publications.

Hruschka, D., Munira, S., Jesmin, K., Hackman, J., & Tiokhin, L. (2018). Learning from Failures of Protocol in Cross-Cultural Research. *Proceedings of the National Academy of Sciences, 115*(45), 11428-11434.

Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence*: Cambridge University Press.

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of international business studies, 19*(3), 411-432.

Laland, K. N. (2017). *Darwin's unfinished symphony: how culture made the human mind*: Princeton University Press.

Mayer, T., & Zignago, S. (2011). Notes on CEPII's distances measures: The GeoDist database.

McCrae, R. R., Terracciano, A., & 79 Members of the Personality Profiles of Cultures Project. (2005). Personality profiles of cultures: aggregate personality traits. *Journal of Personality and Social Psychology, 89*(3), 407.

Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics, 93*(2), 351-363.

Muthukrishna, M., Francois, P., Pourahmadi, S., & Henrich, J. (2017). Corrupting cooperation and how anti-corruption strategies may backfire. *Nature Human Behaviour, 1*(7), 0138.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*.

Muthukrishna, M., Morgan, T. J. H., & Henrich, J. (2016). The When and Who of Social Learning and Conformist Transmission. *Evolution and Human Behavior, 37*(1), 10-20.

Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*: Westview Press.

Pemberton, T. J., DeGiorgio, M., & Rosenberg, N. A. (2013). Population structure in a comprehensive genomic data set on human microsatellite variation. *G3: Genes, Genomes, Genetics, 3*(5), 891-907.

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401-11405.

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America, 102*(44), 15942-15947.

Saucier, G., Kenner, J., Iurino, K., Bou Malham, P., Chen, Z., Thalmayer, A. G., . . . Metaferia, H. (2015). Cross-cultural differences in a global "survey of world views". *Journal of Cross-Cultural Psychology, 46*(1), 53-70.

Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science, 366*(6466), eaau5141.

Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative sociology, 5*(2), 137-182.

Spolaore, E., & Wacziarg, R. (2016). Ancestry and development: New evidence. *Discussion Papers Series, Department of Economics, Tufts University, 820*.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science, 344*(6184), 603-608.

Uz, I. (2015). The Index of Cultural Tightness and Looseness Among 68 Countries. *Journal of Cross-Cultural Psychology, 46*(3), 319-335.

Woodard, C. (2011). *American nations: A history of the eleven rival regional cultures of North America*: Penguin.