Research Article

Psychological Science
2020, Vol. 31(6) 678–701

DOI: 10.1177/0956797620916782
www.psychologicalscience.org/PS

Beyond Western, Educated, Industrial, Rich, and Democratic (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], and Braden Thue[3]

[1]Department of Psychological and Behavioural Science, London School of Economics and Political Science; [2]Department of Anthropology, The University of Utah; [3]Department of Human Evolutionary Biology, Harvard University; and [4]Department of Computer Science, Iowa State University

Abstract

In this article, we present a tool and a method for measuring the psychological and cultural distance between societies and creating a distance scale with any population as the point of comparison. Because psychological data are dominated by samples drawn from Western, educated, industrialized, rich, and democratic (WEIRD) nations, and overwhelmingly, the United States, we focused on distance from the United States. We also present distance from China, the country with the largest population and second largest economy, which is a common cultural comparison. We applied 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, our 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

Received 9/13/18; Revision accepted 1/19/20

Decades of psychological research designed to uncover truths about human psychology may have instead uncovered truths about a thin slice of our species—people who live in Western, educated, industrialized, rich, and democratic (WEIRD) nations (Henrich, Heine, & Norenzayan, 2010). Researchers often assess the generalizability of these findings by comparing Western nations with 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, nondemocratic, 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

Corresponding Author:
Michael Muthukrishna, London School of Economics and Political Science, Department of Psychological and Behavioural Science, Houghton St., London WC2A 2AE, United Kingdom
E-mail: m.muthukrishna@lse.ac.uk

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 with each other and with 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. Because psychological data remain largely American (Rad et al., 2018), we developed an American scale of cultural distance from the United States as an example. As a point of comparison, we also developed a Chinese 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 (2001) cultural dimensions, Inglehart and Welzel's (2005) cultural map, and Schwartz's (2006) values. 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, 2018).

These approaches are widely used but have various limitations. For example, the values-and-dimensions approach characteristic of Hofstede (2001), Inglehart and Welzel (2005), and Schwartz (2006) focuses on identifying the values or dimensions along which groups differ, focusing on cultural differences 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, respectively). Indeed, Brazil and Turkey look very similar along most Hofstede dimensions (see Fig. 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—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 captures the relative frequencies.

Here, we present the cultural fixation index (CF_{ST}), a measure built on the fixation index (F_{ST}) measure from population biology (Bell, Richerson, & McElreath, 2009; Cavalli-Sforza, Menozzi, & Piazza, 1994), as a tool for measuring cultural distance. The CF_{ST} is robust and theoretically defensible. It can be used as a high-resolution method to identify regional, national, or arbitrary cultural groupings (such as class differences). It can also be used to identify the distance between two groups on the basis of 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 (www.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 can be found in the Supplemental Material available online.

The 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

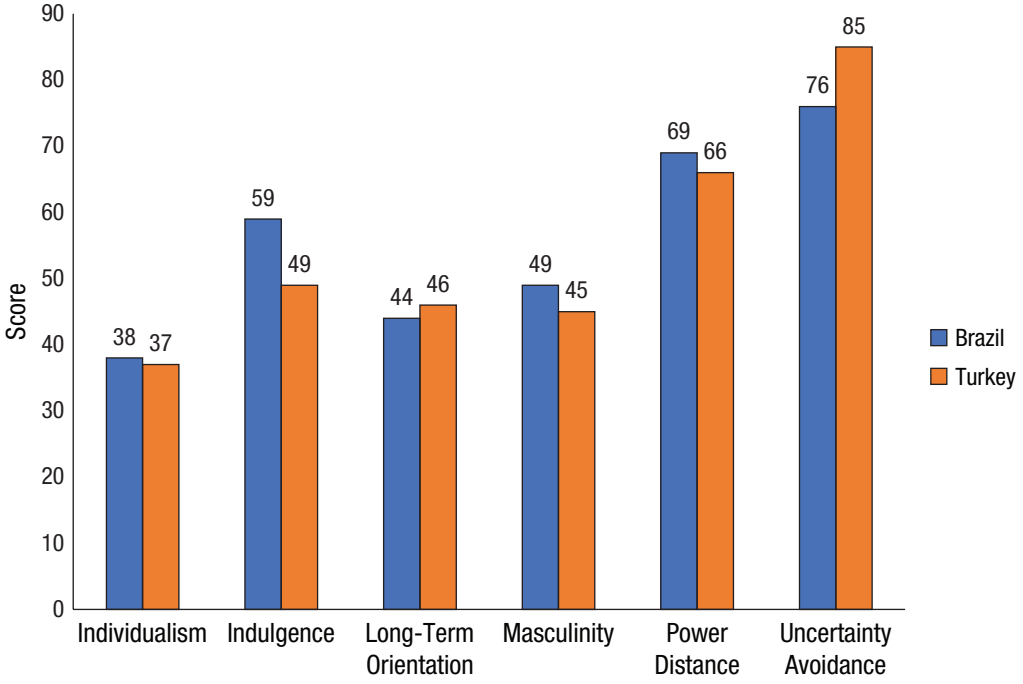


Fig. 1. Brazil's and Turkey's scores on Hofstede's cultural dimensions (data taken from https://www.hofstede-insights.com/product/compare-countries/).

detail, and compare the F_{ST} with other methods, in the Supplemental Material.

We begin by comparing the CF_{ST} with other common approaches to measuring cultural distance. We then formally describe the CF_{ST} statistic and how we used it to develop the American and Chinese 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 we discuss tests of how well they predict common or high-profile measures of cross-cultural psychological differences.

Method

One approach to measuring cultural distance is to turn multiple 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):

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

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, and values and beliefs. This is similar to how any psychometric scale is developed 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 the knowledge of participants from country A about participants from country B. Here, we use the World Values Survey (WVS) as a measure of what people report that they do and believe.

Genetic distance and linguistic distance do not share these limitations; instead, they serve as an overall proxy measure of cultural distance between any two nations. But because they do not measure culture directly, they

can be misleading. For example, Hong Kong is greater than an order of magnitude more genetically similar to China (F_{ST} = 9.59 × 10^{-4}) than to Great Britain (F_{ST} = 3.96 × 10^{-2}; Pemberton, DeGiorgio, & Rosenberg, 2013; Spolaore & Wacziarg, 2018) but is culturally similar to both countries because of 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 United Kingdom, and the United States, all of whose populations 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})

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-groups 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} is equal to $\frac{\sigma_g^2}{\sigma_T^2}$, 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 $\frac{H_T - H_g}{H_T}$, 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. If these pairwise differences are the same, F_{ST} = 0, the populations are identical. If the two equal-size populations are more homogeneous but different, we get the maximum distance, F_{ST} = 1 (see Fig. 2). Like a correlation coefficient, F_{ST} is also a nonlinear metric. This means that the numbers are meaningful for comparisons, but an F_{ST} of .6 is not twice as large as an F_{ST} of .3—just as a correlation (*r*) of .6 is not twice as high a correlation as .3.

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

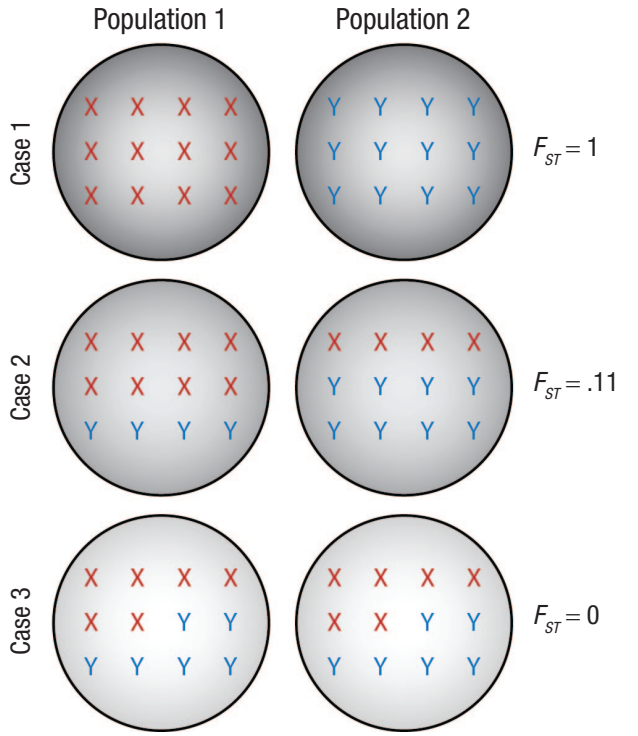


Fig. 2. Three scores on the fixation index (F_{ST}) calculated along a single dimension with two options—yes (Y) or no (X). In Case 3, the cultural distance between the two populations is 0 because 50% of both populations answered yes (or no). In Case 2, the F_{ST} is .11. Finally, in Case 1, the populations are maximally different because all individuals in Population 1 answered no, and all individuals in Population 2 answered yes. Calculations for each of these cases can be found in the Supplemental Material available online. For the two populations, we calculated the mean F_{ST} along all cultural questions or along specific questions of interest (such as those mapping onto a particular dimension).

population. The availability of large, representative, cross-national surveys of cultural values has allowed the same technique to be applied to culture.

Summary of the CF_{ST}

The CF_{ST} is calculated in the same manner as the 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 that, 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. The CF_{ST} allows us to flexibly quantify the overall size of cultural differences (cultural distance). In contrast to many other

measures of cultural differences, the 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, the CF_{ST} can handle continuous, binary, or nominal traits. By calculating CF_{ST} scores 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 toward their children, and the various values that parents wish to instill in their children. These dimensions may be predetermined (similar to including all genes associated with height or skin color) or based on dimension-reduction techniques. A list of predetermined dimensions and statistically reduced dimensions can be found in the Data section. Researchers can examine predetermined dimensions at www.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 the CF_{ST}. The CF_{ST} can also be calculated between regions within larger nations; nations are not equally heterogeneous. Such within-countries 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 or questions within subdimensions (see examples in the Supplemental Material). Moreover, we can calculate confidence intervals (CIs) as a measure of uncertainty.

Formal description of the $\mathbf{CF_{ST}}$

The CF_{ST} can be computed for any measured trait as long as there are 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) or categorical (nominal). The 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 the CF_{ST}, we might be interested only 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 because we have no particular theoretical basis for examining subdimensions, here we focus on overall cultural distance as a demonstration. We 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, the 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 the data from 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 $\frac{\sum_{i=1}^{s}(x_i - \bar{x})^2}{s - 1}$, 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:

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

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

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

Data

Cultural distance is calculated using combined data from the two most recent waves of the WVS (2005–2009 and 2010–2014; Inglehart et al., 2014). These waves contain answers from 170,247 participants gathered from nationally representative samples of 80 countries where approximately 85% of the world's population lives. 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); Bell and colleagues compared only neighboring countries within the same region, so were able to include these. We also excluded all demographic questions (e.g., age, sex). A full list of our inclusion and exclusion decisions for all WVS questions is available in the Supplemental Material.

We used the CF_{ST} to develop an American scale of cultural distance from the United States and a Chinese scale of cultural distance from China. We compared these scales with the following cultural differences, psychological outcomes, and distance measures.

Hofstede's cultural dimensions. Hofstede's (2001) cultural dimensions 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 (n.d.) 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 (such as the family) accept and expect that power is distributed unequally. Masculinity is the extent to which the use of force [is] 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-[term]-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. (https://geerthofstede.com/culture-geert-hofstede-gert-jan-hofstede/6d-model-of-national-culture/)

Tightness–looseness. Gelfand et al. (2011) described tight societies as having many strong norms and a low tolerance for deviant behavior. In contrast, loose societies are characterized as having weak social norms and a high tolerance for deviant behavior. Gelfand et al. measured tightness on the basis of 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 on the basis of its outcome—greater variance or standard deviation in professed cultural values. Uz introduced three standard-deviation measures of answers in the 2000 wave of the WVS. These included (a) a domain-specific index of the standard deviation of answers to the Morally Debatable Behaviors Scale, which included attitudes toward prostitution, abortion, divorce, euthanasia, and suicide; (b) a domain-general index based on the standard deviation of a variety of values and behavioral practices in the WVS; and (c) a composite measure based on a factor analysis of the domains in the WVS. This combined or composite measure had the greatest validity. All Uz measures of looseness are correlated only weakly with the Gelfand et al. (2011) measure (rs = .16–.30).

Schwartz's values. Schwartz's (2006) cultural-value orientation is a theory that outlines various values that help regulate human behavior in different societies. *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 used 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 133 to 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 1,000 individuals were used as a measure of impersonal cooperative altruism. The frequency of voluntary, unpaid, anonymous blood donations from countries (World Health Organization) and Italian provinces was measured. These data were collated by Schulz et al. (2019) from the World Health Organization *Global Status Report on Blood Safety and Availability*. Diplomat Parking Tickets is based on data on unpaid parking tickets of United Nations diplomats in New York City. These data come from a classic economics study showing that the number of unpaid parking tickets is predicted by corruption norms (Fisman & Miguel, 2007). Diplomats do not face sanctions for unpaid parking tickets because of 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 (2016), using the 2015 index to match the WVS. In addition to predicting the accrual of diplomats' parking tickets, it also predicts bribery behavior within economic games (Muthukrishna, Francois, Pourahmadi, & Henrich, 2017). We used the 2015 index, the final year of the current wave of the WVS. The Return Wallet data are from a recent study comparing return rates of dropped wallets in several countries around the world (Cohn, Maréchal, Tannenbaum, & Zünd, 2019).

Distance measures. The Centre d'Études Prospectives et d'Informations Internationales (CEPII) GeoDist data set (Mayer & Zignago, 2012) contains pairwise geographic distance measures between countries calculated in four different ways. The first measure (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 measure (Geographic Distance Capitals) is calculated using 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 data set (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 (www.ethnologue.com). Linguistic Distance ASJP is based on the results of the Automated Similarity Judgment Program (ASJP) project (https://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.'s (2013) work, matched to countries by Spolaore and Wacziarg (2018). Genetic Distance Ethnic Weighting weights the ethnic groups within a country by their population size. Genetic Distance Ethnic Plurality considers only genetic distance from the largest ethnic plurality in the country.

Answers as alleles. The WVS contains various answer types, including Likert-type scales, binary approve/disapprove responses, and categorical responses. Societies have markedly different answering styles (Chen, Lee, & Stevenson, 1995; Heine, Lehman, Peng, & Greenholtz, 2002). There are also issues with using Likert-type scales cross-culturally for a variety of reasons, including cross-cultural differences in preferences for yes/no answers versus a Likert-type scale rating (for discussion and a review, see Hruschka, Munira, Jesmin, Hackman, & 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 midpoint 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 showing only opposite beliefs or behaviors. That is, larger distances may exist because of 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 cannot 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 responses 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 questions had very few responses for a particular option.

Our decision strategy was as follows. First, we changed all valence questions to a positive allele and a negative allele. A midpoint was coded as a third allele. For example, answers options for the question "Indicate how important it is in your life: Family" were *very important, rather important, not very important*, and *not at all important*. These were condensed to two cultural alleles: Allele 1 consisted of *very important* and *rather important*, and Allele 2 consisted of *not very important* and *not at all important*. Another example is the question, "To fully develop your talents, you need to have a job," which had answer options of *strongly agree, agree, neither, disagree*, and *strongly disagree*. These were condensed to three alleles—one for each valence and one for the midpoint: Allele 1 consisted of *strongly agree* and *agree*, Allele 2 consisted of *disagree* and *strongly disagree*, and Allele 3 consisted of *neither*.

Second, we restricted this breakdown to a maximum of four alleles, combining multiple answers on the basis of overall distributions of responses. For examples, the answer options for the question "Apart from weddings and funerals, about how often do you attend religious services these days?" were *more than once a week, once a week, once a month, only holy days, once a year, less often*, and *never* (the wording was changed slightly between waves). These were condensed to four alleles: Allele 1 consisted of *more than once a week* and *once a week*, Allele 2 consisted of *once a month* and *only holy days*, Allele 3 consisted of *once a year* and *less often*, and Allele 4 consisted of *never*. (A full list of allele categorizations is available in the file "allele-dimension-data.csv" in the Supplemental Material.)

CIs and robustness calculations

We calculated 95% CIs by bootstrapping with 1,000 replications as recommended by Bell et al. (2009). Figures S1 and S2 in the Supplemental Material show the American and Chinese scales, respectively, with 95% CIs.

We investigated the robustness of the CF_{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 compared the calculated CF_{ST} scores with the scores using the full data set. We did this 10,000 times for each percentage value for each method of resampling and then compared the result with the real CF_{ST} values, recording (a) the size of the deviation, (b) the Pearson correlation, and (c) 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 Chinese scale with all countries relative to China.

Results

American scale and Chinese 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 do not unidimensionally range from WEIRD to non-WEIRD, we also constructed a Chinese scale by calculating the cultural distance for all countries relative to China, a common cultural comparison in cultural psychology.

Because the CF_{ST} is a composite of many questions, we effectively cut a line through a large multidimensional culture space and looked only at the distance from a particular point (the United States in the American scale, China in the Chinese scale). Thus, Japan and Norway are similarly culturally distant from the United States (.115 and .124, respectively) 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 Chinese scale of cultural distance, which are also graphed on a map in Figure 3 and on number lines in Figure 4 (note that the most commonly studied non-Western nations—Japan, Hong Kong, and China—are by no means the extreme on the American scale). Figures S1 and S2 show the American scale and Chinese scale number lines, respectively, with 95% CIs shaded in blue. Further emphasizing that cultural distances are distances in a large multidimensional culture space, Figure 5 shows a 2-D 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 (and are not necessarily culturally close to each other). For example, Sweden and Colombia are similarly distant from both the United States (CF_{ST} scores = .115 and .102, respectively) and China (CF_{ST} scores = .186

Table 1. Values From the American Scale and Chinese Scale of Cultural Distance

	American		Chinese	
Country	Cultural distance	95% CI	Cultural distance	95% CI
Algeria	.138	[.132, .144]	.221	[.216, .228]
Andorra	.115	[.109, .122]	.249	[.242, .258]
Argentina	.071	[.069, .075]	.150	[.146, .155]
Armenia	.154	[.149, .161]	.177	[.171, .183]
Australia	.035	[.033, .039]	.131	[.127, .135]
Azerbaijan	.175	[.169, .181]	.158	[.153, .165]
Bahrain	.167	[.161, .173]	.195	[.189, .201]
Belarus	.071	[.068, .075]	.101	[.097, .106]
Brazil	.072	[.069, .075]	.159	[.156, .163]
Bulgaria	.108	[.104, .114]	.116	[.111, .123]
Burkina Faso	.143	[.139, .149]	.153	[.149, .157]
Canada	.026	[.025, .028]	.135	[.132, .140]
Chile	.078	[.075, .081]	.156	[.152, .161]
China	.150	[.146, .155]	—	—
Colombia	.102	[.0987, .106]	.182	[.178, .186]
Cyprus	.057	[.055, .061]	.118	[.114, .122]
Ecuador	.109	[.105, .114]	.197	[.192, .204]
Egypt	.234	[.228, .241]	.186	[.183, .190]
Estonia	.117	[.112, .122]	.097	[.093, .102]
Ethiopia	.130	[.126, .136]	.153	[.149, .158]
Finland	.072	[.069, .076]	.176	[.171, .185]
France	.079	[.075, .085]	.181	[.175, .190]
Georgia	.143	[.139, .148]	.143	[.140, .146]
Germany	.080	[.078, .084]	.114	[.111, .117]
Ghana	.153	[.149, .158]	.172	[.169, .175]
Great Britain	.046	[.043, .051]	.172	[.166, .181]
Guatemala	.134	[.130, .140]	.192	[.186, .198]
Hong Kong	.090	[.088, .095]	.085	[.082, .090]
Hungary	.102	[.098, .108]	.125	[.121, .132]
India	.093	[.091, .097]	.106	[.104, .110]
Indonesia	.178	[.173, .184]	.167	[.163, .171]
Iran	.150	[.145, .156]	.125	[.122, .128]
Iraq	.162	[.158, .167]	.193	[.189, .197]
Italy	.061	[.059, .065]	.163	[.157, .169]
Japan	.115	[.112, .119]	.118	[.114, .122]
Jordan	.195	[.190, .200]	.193	[.189, .197]
Kazakhstan	.107	[.103, .111]	.099	[.095, .104]
Kuwait	.122	[.117, .127]	.163	[.157, .169]
Kyrgyzstan	.132	[.128, .137]	.161	[.156, .166]
Lebanon	.103	[.099, .109]	.175	[.169, .182]
Libya	.146	[.142, .151]	.198	[.194, .202]
Malaysia	.125	[.121, .129]	.156	[.153, .160]
Mali	.155	[.150, .161]	.155	[.151, .160]
Mexico	.077	[.074, .080]	.138	[.135, .141]
Moldova	.100	[.096, .105]	.133	[.128, .140]
Morocco	.149	[.145, .155]	.176	[.172, .180]
Netherlands	.079	[.076, .083]	.146	[.142, .150]
New Zealand	.053	[.050, .058]	.162	[.156, .168]
Nigeria	.130	[.126, .135]	.222	[.217, .227]

(continued)

Table 1. *(continued)*

	American		Chinese	
Country	Cultural distance	95% CI	Cultural distance	95% CI
Norway	.124	[.118, .131]	.206	[.199, .214]
Pakistan	.178	[.173, .185]	.240	[.234, .246]
Palestine	.134	[.129, .140]	.193	[.187, .20]
Peru	.090	[.087, .094]	.142	[.139, .146]
Philippines	.144	[.139, .150]	.229	[.223, .236]
Poland	.076	[.074, .081]	.147	[.143, .151]
Qatar	.176	[.171, .183]	.262	[.255, .269]
Romania	.103	[.100, .108]	.140	[.137, .144]
Russia	.085	[.083, .088]	.089	[.086, .092]
Rwanda	.149	[.145, .154]	.143	[.140, .146]
Serbia and Montenegro[a]	.079	[.076, .084]	.166	[.160, .174]
Singapore	.038	[.036, .041]	.124	[.120, .129]
Slovenia	.077	[.074, .081]	.122	[.118, .126]
South Africa	.076	[.073, .079]	.138	[.135, .141]
South Korea	.092	[.089, .095]	.073	[.071, .077]
Spain	.074	[.071, .078]	.137	[.133, .142]
Sweden	.115	[.111, .121]	.186	[.180, .191]
Switzerland	.068	[.064, .074]	.179	[.173, .187]
Taiwan	.097	[.095, .101]	.092	[.089, .096]
Thailand	.129	[.125, .134]	.104	[.101, .107]
Trinidad and Tobago	.088	[.085, .093]	.187	[.183, .191]
Tunisia	.156	[.151, .163]	.179	[.175, .185]
Turkey	.120	[.117, .126]	.119	[.117, .122]
Ukraine	.086	[.083, .089]	.117	[.114, .123]
United States	—	—	.150	[.146, .155]
Uruguay	.084	[.081, .088]	.143	[.139, .148]
Uzbekistan	.150	[.146, .155]	.150	[.146, .155]
Viet Nam	.182	[.177, .188]	.057	[.055, .061]
Yemen	.200	[.193, .209]	.248	[.241, .256]
Zambia	.083	[.081, .088]	.162	[.158, .167]
Zimbabwe	.110	[.106, .115]	.220	[.216, .226]

Note: Cultural distances are differences in the cultural fixation index (CF_{ST}) between a given country and (a) the United States and (b) China. CI = confidence interval.
[a]Serbia and Montenegro separated in 2006; however, they remain combined in the World Values Survey data.

and .182, respectively), but they are not culturally close to each other (indeed, they are very different; CF_{ST} score = .261). 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 Figs. 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: $r = .86$, $p < .001$; and British scale: $r = .83$, $p < .001$). Whenever 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 U.S. studies will be less powerful if done in Australia compared with Yemen, but generalizing a study conducted in Turkey will be less powerful if done in Yemen compared with Australia.

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

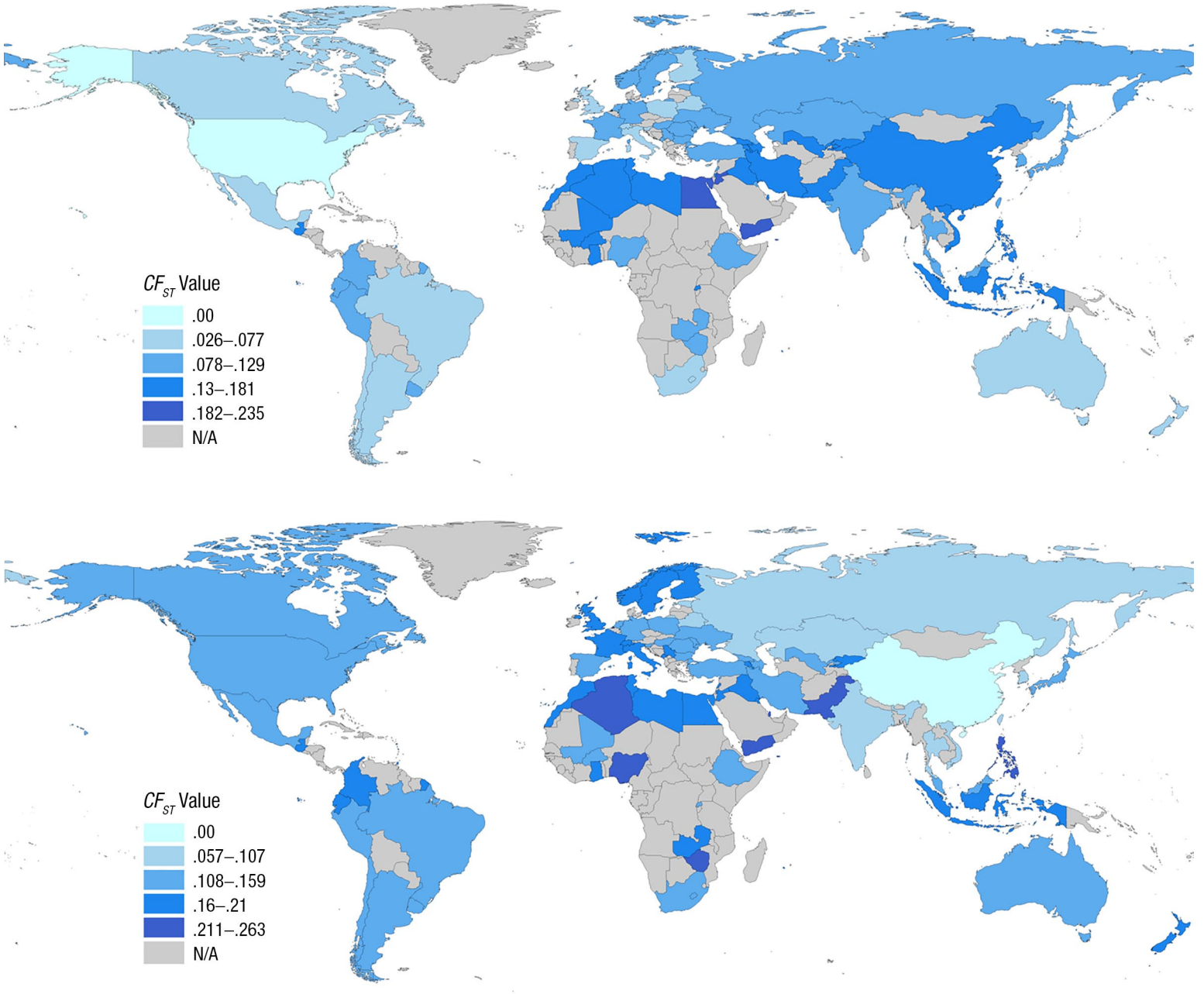


Fig. 3. Cultural distances on the cultural fixation index (CF_{ST}) visualized on a world map. The top map shows the distance of countries from the United States on the American scale of cultural distance. The bottom map shows the distance of countries from China on the Chinese scale of cultural distance.

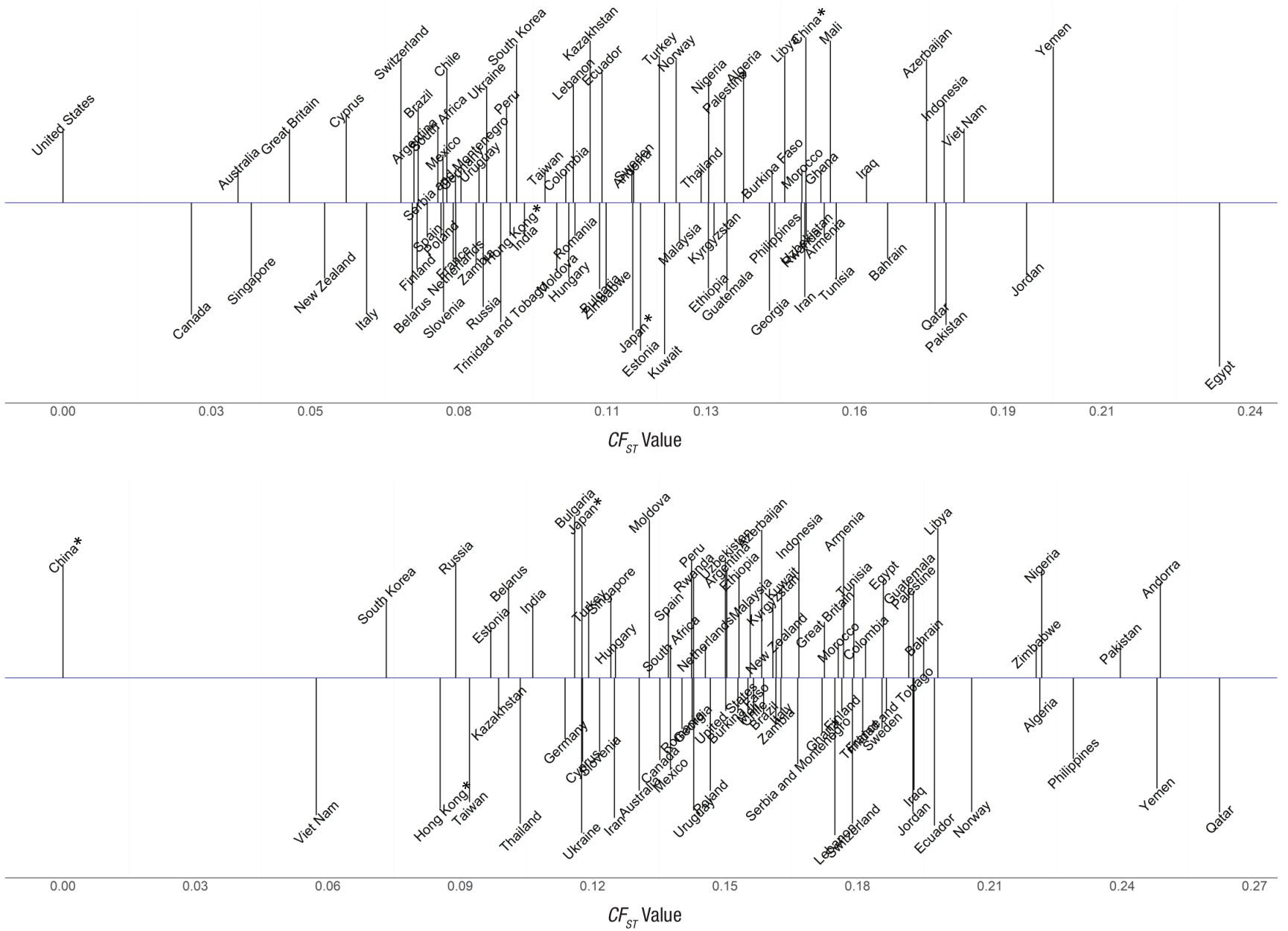


Fig. 4. Cultural distances on the cultural fixation index (CF_{ST}) visualized on a number line. The top graph shows the distance of countries from the United States on the American scale of cultural distance. The bottom graph shows the distance of countries from China on the Chinese scale of cultural distance. The most commonly studied non-Western nations are marked with asterisks.

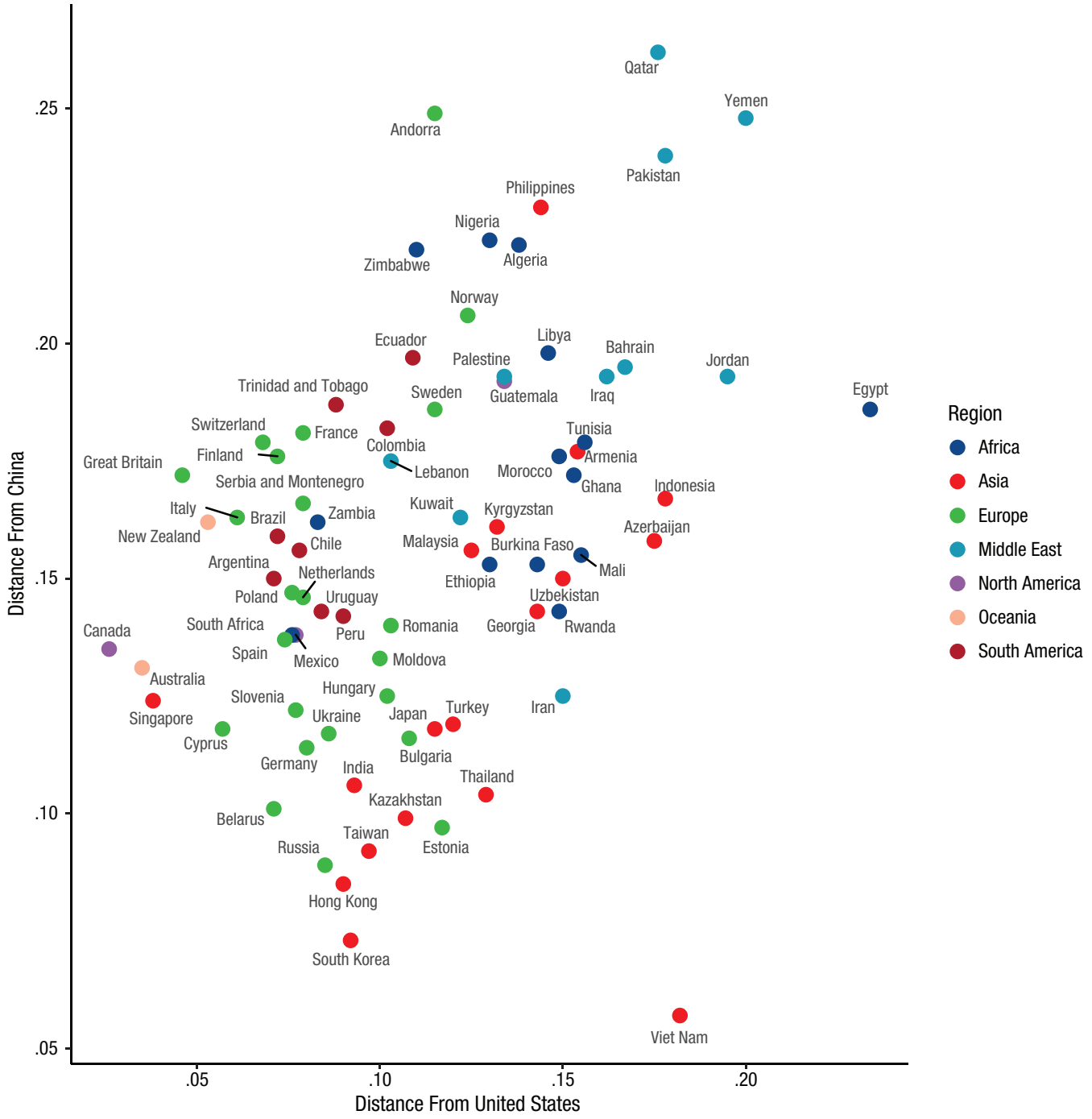


Fig. 5. Scatterplot showing the relation between countries' scores on the American and Chinese cultural fixation index (CF_{ST}), color-coded by region.

the R code and created an online tool (www.culturaldistance.com). In the future, this tool will also allow region-by-region comparisons.

Within-nations 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 reveal the relative homogeneity of the United States compared with these other large populations.

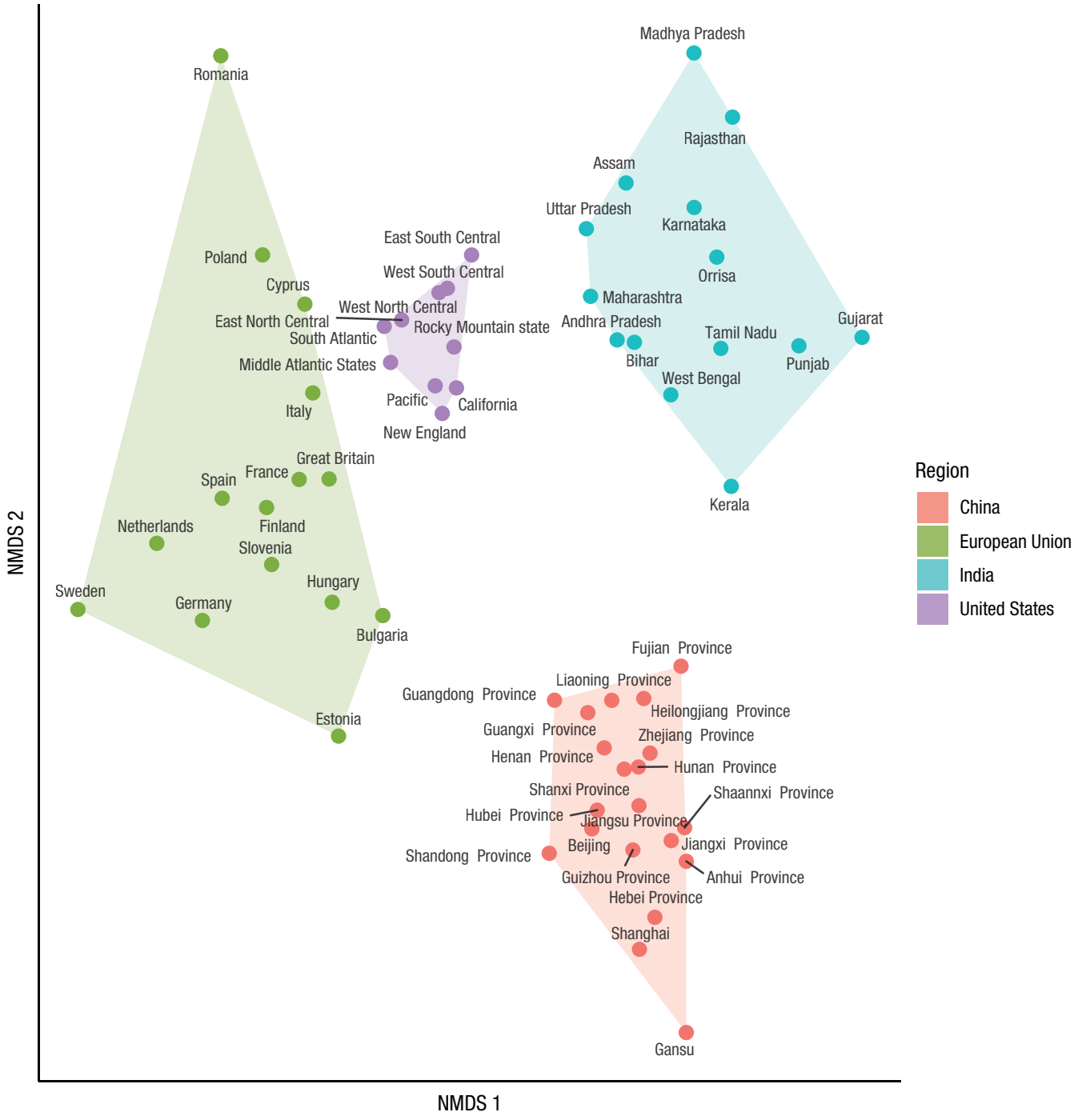


Fig. 6. Two-dimensional nonmetric multidimensional scaling (NMDS) plot of the pairwise cultural fixation index (CF_{ST}) matrix of regions with at least 100 survey participants in the four largest populations: United States, China, India, and the countries of the European Union.

India, "the country of a hundred nations and a hundred tongues, of a thousand religions and two million gods" (Twain, 1897/1989, p. 348), has the largest mean regional diversity of these populations (CF_{ST} = .11). The European Union, a long-running project attempting to create a political and economic union, is next (CF_{ST} = .09), followed by China, despite being 90% Han (CF_{ST} = .05). The United States has the least regional diversity with a CF_{ST} of .01.

We illustrate these differences in Figure 6, displaying nonmetric multidimensional scaling with the *R MASS* package (Venables & Ripley, 2002) on a pairwise CF_{ST} matrix including the WVS regions within the United

States, China, India, and the European Union with at least 100 surveyed individuals.

These within-population analyses reveal the importance of remembering that societies are not homogeneous but, rather, have multivariate distributions of many traits along many dimensions with structure within structure. There are likely to be cultural differences between not only regions within a country but also ethnicities, religions, socioeconomic class, and other groupings. These are all avenues for future research.

Robustness tests

We investigated the robustness of the CF_{ST} by randomly resampling a fixed proportion of the question set and comparing these values with the full question set. We did this 10,000 times for each proportion value and then (a) counted the number of times a value fell outside the CI 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 Chinese scale, we did the same with all countries relative to China. We plot these in Figures 7 and 8.

These analyses show that the CF_{ST} is highly robust. We can sample up to only 50% of questions and still get only small deviations (.013 and .026 from the United States and China, respectively) and values that remain highly correlated with the original scale by both Pearson correlation (rs = .95 and .87 from the United States and China, respectively) and Spearman correlation (rs = .95 and .85 from the United States and China, respectively). When we sample 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) because of 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, the CF_{ST} will be robust even if we do not 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 with other measured psychological and cultural differences

In Table 2, we compare the American scale and Chinese scale with 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 (labeled "relative"). For the distance scores, we compare the distance from the United States and China, respectively. We plot these correlations visually in Figure S4 in the Supplemental Material.

These correlations reveal that the American scale predicts many more cross-cultural differences in psychology than the Chinese scale. Remarkably, many of these correlations are large and significant even when we correlated 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 nonexclusive possibility is that researchers from WEIRD countries have focused their data-gathering efforts on psychological dimensions that they find odd in comparison with Western psychology. Long-Term Orientation, a new addition to Hofstede's cultural dimensions, is a nice illustration of a trait for which China, rather than the United States, is the outlier. If psychology were dominated by China, Table 2 would have perhaps been dominated by dimensions that Chinese researchers found to be different in other parts of the world, and the Chinese scale, rather than the American scale, would consequently be more predictive of differences.

It is reassuring that, in contrast to these psychological differences, the proxies for cultural distance—geographic, linguistic, and genetic—although 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 (a) we are measuring an overall cultural distance—a strength of the CF_{ST} approach, (b) an American scale is predictive of various psychological variables, and (c) the United States is unique either as a psychological outlier or because of what the United States and other WEIRD researchers have chosen to study.

Hofstede's (2001) 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, for which societies that have more emphasis on individual freedom are more culturally similar to the United States. These correlations match U.S. stereotypes that emphasize equality and individual

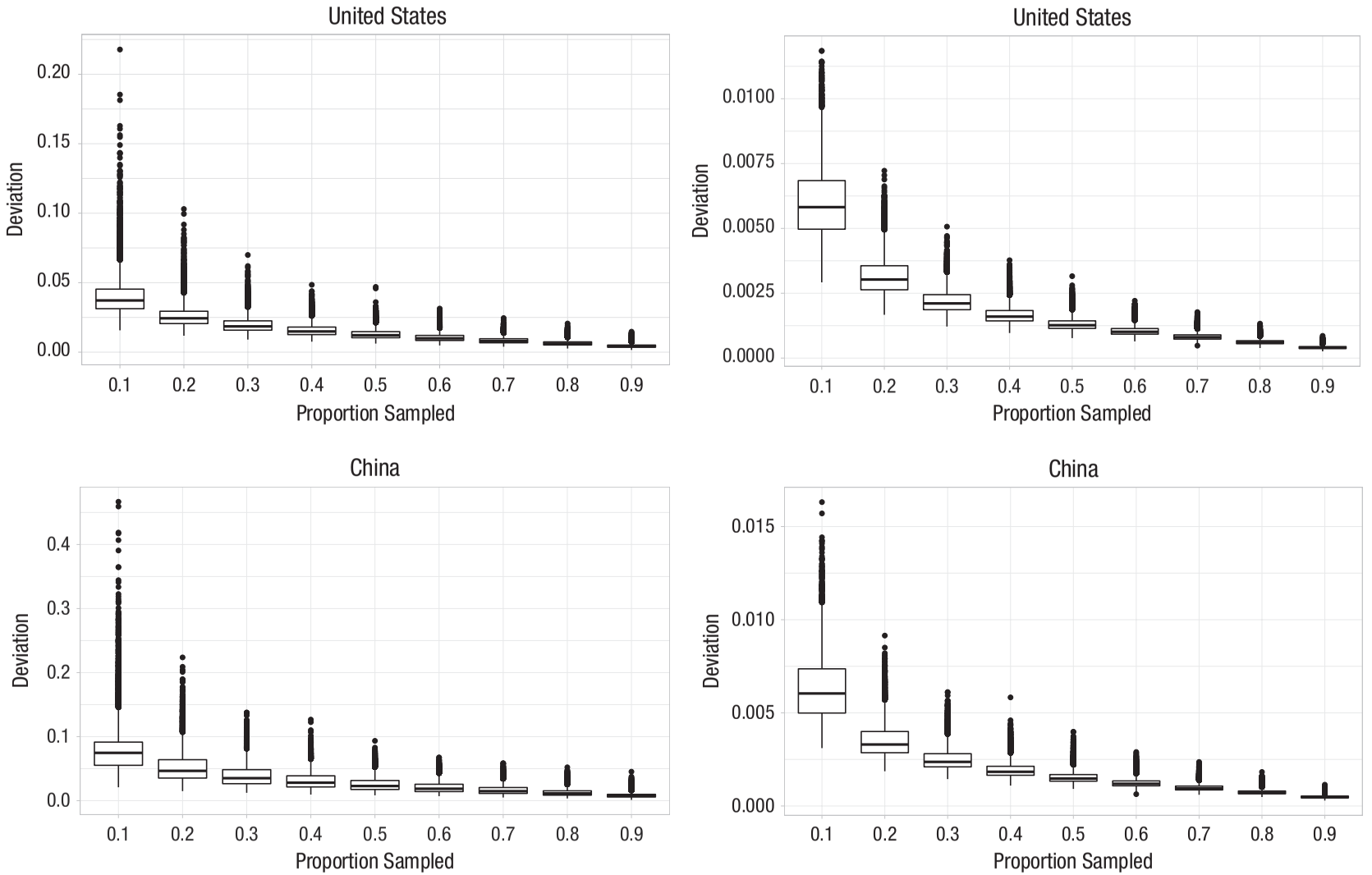
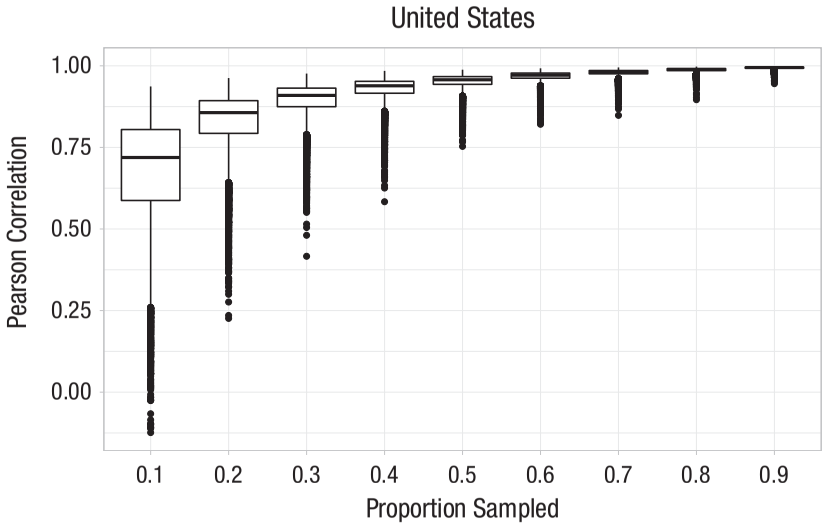
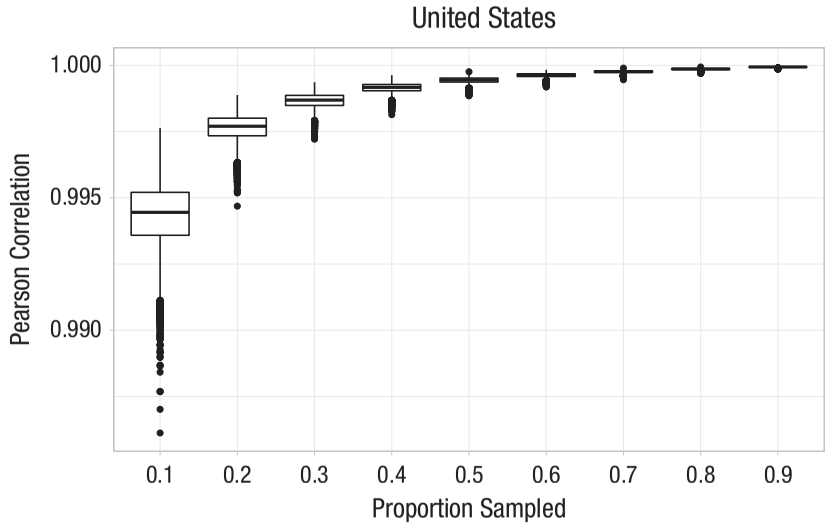
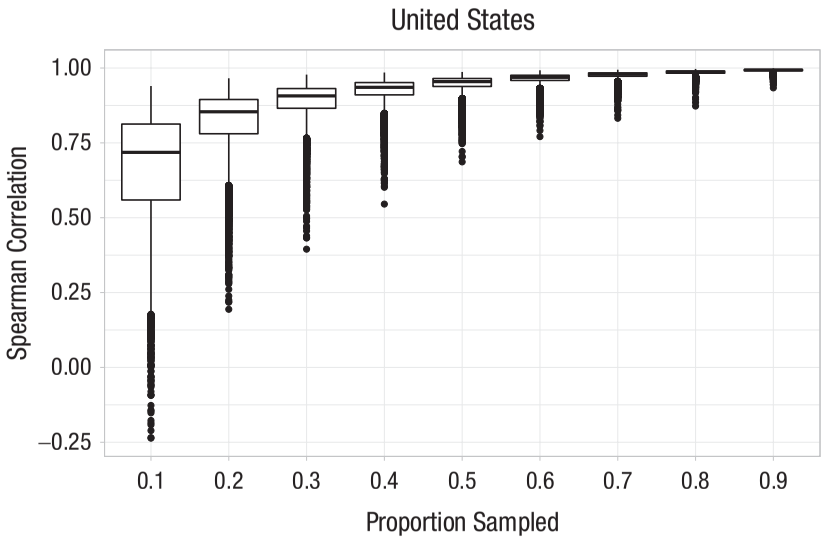


Fig. 7. Size of the deviations in the 10,000 resampled values for each proportion sampled, separately for the American (top row) and Chinese (bottom row) scales. The graphs on the left show sampling by the proportion of questions. The graphs on the right show sampling by the proportion of values. In each box-and-whisker plot, the central horizontal line indicates the median; the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively; whiskers mark 1.5 times the interquartile range from the 25th and 75th percentiles; and dots indicate outlier data points.







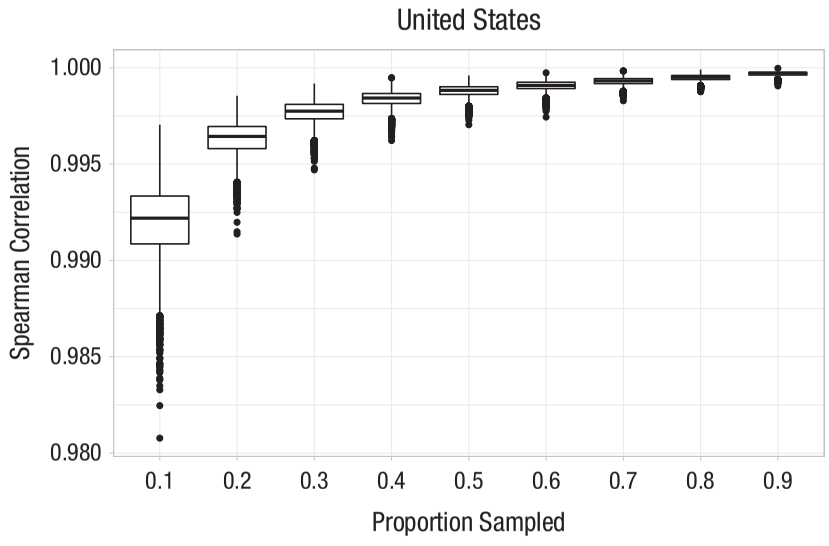


Fig. 8. *(continued on next page)*

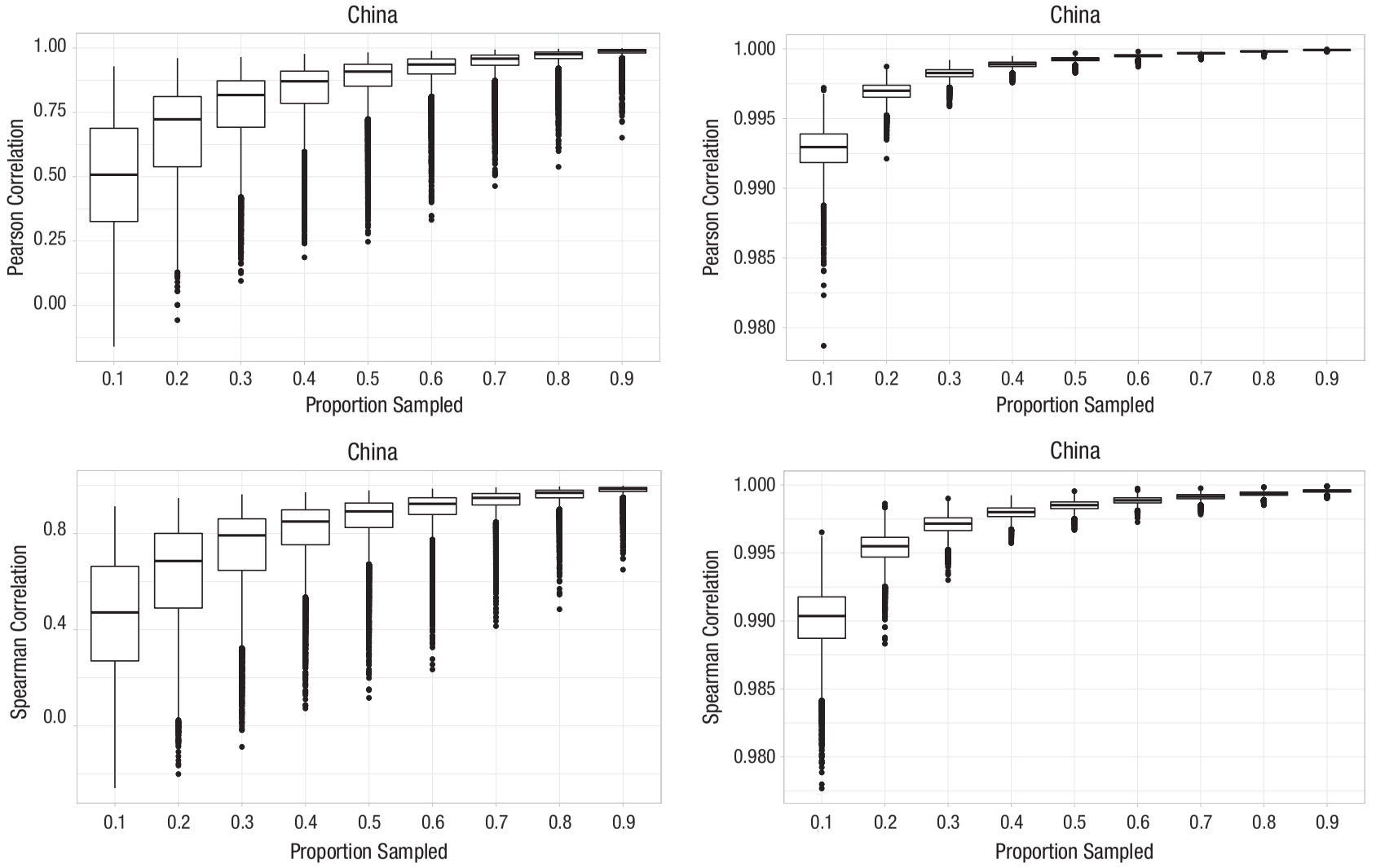


Fig. 8. Spearman and Pearson correlations in the 10,000 resampled values for each proportion sampled, separately for the American (top two rows) and Chinese (bottom two rows) scales. The graphs on the left show sampling by the proportion of questions. The graphs on the right show sampling by the proportion of values. In each box-and-whisker plot, the central horizontal line indicates the median; the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively; whiskers mark 1.5 times the interquartile range from the 25th and 75th percentiles; and dots indicate outlier data points.

Table 2. Correlations Between the American Scale (CF_{ST} Distance From the United States), the Chinese Scale (CF_{ST} Distance From China), and Other Commonly Used Psychological and Cultural Difference and Distance Measures

	American scale			Chinese scale		
Correlated measure	*r*	95% CI	*df*	*r*	95% CI	*df*
Hofstede's cultural dimensions						
Individualism[a]	−.51***	[−.68, −.30]	57	−.02	[−.27, .24]	57
Individualism relative[a]	.51***	[.30, .68]	57	.06	[−.20, .31]	57
Power Distance[a]	.42***	[.19, .61]	57	.06	[−.20, .31]	57
Power Distance relative[a]	.40**	[.16, .60]	57	.00	[−.25, .26]	57
Masculinity[a]	−.06	[−.32, .19]	57	.07	[−.19, .32]	57
Masculinity relative[a]	.06	[−.20, .31]	57	.20	[−.06, .44]	57
Uncertainty Avoidance[a]	.00	[−.25, .26]	57	−.03	[−.29, .22]	57
Uncertainty Avoidance relative[a]	−.06	[−.31, .20]	57	−.29*	[−.50, −.03]	57
Long-Term Orientation[a]	−.23†	[−.46, .04]	53	−.55***	[−.71, −.34]	53
Long-Term Orientation relative[a]	−.06	[−.32, .21]	53	.53***	[.31, .70]	53
Indulgence[a]	−.44***	[−.63, −.20]	53	.20	[−.07, .44]	53
Indulgence relative[a]	.50***	[.27, .67]	53	.33*	[.07, .55]	53
Tightness–looseness						
Tightness Gelfand[b]	.41*	[.02, .70]	23	.11	[−.30, .48]	23
Tightness Gelfand relative[b]	.62**	[.29, .81]	23	.21	[−.20, .56]	23
Looseness composite[c]	−.75***	[−.86, −.57]	38	−.12	[−.41, .19]	39
Looseness composite relative[c]	.54***	[.28, .73]	38	.18	[−.13, .47]	39
Looseness domain specific[c]	−.71***	[−.83, −.52]	40	−.39*	[−.62, −.10]	40
Looseness domain specific relative[c]	.73***	[.55, .85]	40	.06	[−.25, .36]	40
Looseness domain general[c]	−.81***	[−.90, −.66]	38	−.30†	[−.56, .01]	39
Looseness domain general relative[c]	.69***	[.48, .82]	38	−.11	[−.40, .21]	39
Schwartz's values						
Harmony value[d]	−.25†	[−.49, .03]	50	−.06	[−.32, .22]	50
Harmony cultural relative[d]	−.25†	[−.49, .03]	50	−.09	[−.35, .19]	50
Mastery value[d]	−.06	[−.32, .22]	50	−.21	[−.46, .07]	50
Mastery value relative[d]	.25†	[−.03, .49]	50	.21	[−.07, .46]	50
Embeddedness value[d]	.66***	[.47, .79]	50	.21	[−.06, .46]	50
Embeddedness value relative[d]	.56***	[.33, .72]	50	.58***	[.37, .74]	50
Hierarchy value[d]	.27†	[−.00, .50]	50	−.22	[−.46, .06]	50
Hierarchy value relative[d]	−.03	[−.30, .25]	50	.22	[−.06, .46]	50
Egalitarianism value[d]	−.40**	[−.61, −.15]	50	.27*	[.00, .51]	50
Egalitarianism value relative[d]	−.10	[−.36, .18]	50	.27†	[−.00, .50]	50
Affective autonomy value[d]	−.57***	[−.73, −.35]	50	−.16	[−.42, .11]	50
Affective autonomy value relative[d]	.55***	[.32, .71]	50	.39**	[.13, .60]	50
Intellectual autonomy value[d]	−.49***	[−.67, −.25]	50	−.15	[−.41, .13]	50
Intellectual autonomy value relative[d]	−.16	[−.42, .11]	50	.34**	[.07, .56]	50
Five-factor model of personality						
Openness[e]	−.29†	[−.57, .05]	33	.16	[−.18, .47]	33
Openness relative[e]	−.15	[−.46, .19]	33	−.01	[−.34, .33]	33
Conscientiousness[e]	−.09	[−.41, .25]	33	−.04	[−.37, .29]	33
Conscientiousness relative[e]	.18	[−.16, .49]	33	.20	[−.15, .50]	33
Extraversion[e]	−.53**	[−.73, −.23]	33	−.12	[−.44, .22]	33
Extraversion relative[e]	.48**	[.18, .70]	33	−.00	[−.34, .33]	33
Agreeableness[e]	−.33†	[−.59, .01]	33	−.17	[−.48, .17]	33
Agreeableness relative[e]	−.00	[−.33, .33]	33	.15	[−.19, .46]	33
Neuroticism[e]	−.09	[−.42, .25]	33	.09	[−.25, .41]	33
Neuroticism relative[e]	−.06	[−.39, .28]	33	.09	[−.25, .41]	33
Personality standard deviation[e]	−.40*	[−.64, −.07]	33	−.19	[−.49, .16]	33
Personality standard deviation relative[e]	.53**	[.23, .73]	33	−.01	[−.34, .33]	33

(continued)

Table 2. *(continued)*

Correlated measure	American scale			Chinese scale		
	r	95% CI	*df*	*r*	95% CI	*df*
Other psychological and behavioral measures						
Blood donations[f]	−.50***	[−.66, −.29]	63	−.34*	[−.54, −.11]	63
Blood donations relative[f]	.51***	[.31, .67]	63	−.29*	[−.50, −.05]	63
Diplomat Parking Tickets[g]	.40***	[.19, .58]	67	.14	[−.10, .37]	67
Diplomat Parking Tickets relative[g]			67	.16	[−.08, .39]	67
Corruption CPI[h]	−.50***	[−.65, −.31]	74	−.15	[−.36, .08]	74
Corruption CPI relative[h]	.47***	[.28, −.63]	74	−.03	[−.25, .20]	74
Return wallet without money[i]	−.53**	[−.75, −.21]	31	.32†	[−.05, .61]	31
Return wallet without money relative[i]	.45*	[.10, .70]	31	.32†	[−.05, .61]	31
Return wallet with money[i]	−.49**	[−.72, −.16]	31	.23	[−.15, .54]	31
Return wallet with money relative[i]	.51**	[.18, .73]	31	.23	[−.14, .54]	31
Distance measures						
Kogut-Singh cultural distance original[j]	.41**	[.17, .60]	57	.01	[−.24, .27]	57
Kogut-Singh cultural distance all[j]	.43**	[.18, .62]	57	.37**	[.11, .58]	57
Geographic Distance Population Center[k]	.21†	[−.01, .42]	72	.25*	[.02, .45]	72
Geographic Distance Capitals[k]	.23*	[.00, .44]	72	.25*	[.03, .46]	72
Geographic Distance Gravity Weight 1[k]	.29*	[.06, .48]	72	.26*	[.04, .46]	72
Geographic Distance Gravity Weight 2[k]	.29*	[.07, .49]	72	.27*	[.05, .47]	72
Linguistic Distance Ethnologue[l]	.14	[−.17, .43]	38	—	—	38
Linguistic Distance ASJP[l]	−.17	[−.39, .08]	65	.14	[−.10, .38]	63
Genetic Distance Ethnic Weighting[m]	.21†	[−.02, .42]	72	.37**	[.16, .55]	72
Genetic Distance Ethnic Plurality[m]	.17	[−.07, .38]	72	.38***	[.17, .56]	72

Note: Although low obedience and high creativity were identified by Schulz, Bahrami-Rad, Beauchamp, and Henrich (2019) as part of a Western, educated, industrialized, rich, and democratic package, we did not include them here because they are derived from World Values Survey (WVS) questions. CF_{ST} = cultural fixation index; CI = confidence interval; ASJP = Automated Similarity Judgment Program.

[a]These values were taken 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). [b]These values were taken from Gelfand et al. (2011). Higher scores indicate greater tightness. The mean for East and West Germany was used for Germany. Relative values are absolute values relative to the comparison country (United States or China). [c]These values were taken from Uz (2015), whose measure of looseness uses variance in WVS 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 used the domain-specific value as a proxy. [d]These values are Schwartz's (2006) culture-value orientation scores. The mean for East and West Germany was used for Germany. The mean for French and German Switzerland was used for Switzerland. Relative values are absolute values relative to the comparison country (United States or China). [e]The personality-factor data for each country were taken from Table 2 in the study by McCrae, Terracciano, and 79 Members of the Personality Profiles of Cultures Project (2005). The mean for French and German Switzerland was used for Switzerland. Relative values are absolute values relative to the comparison country (United States or China). [f]Blood-donations data per 1,000 persons were collated from the World Health Organization *Global Status Report on Blood Safety and Availability* (Schulz et al., 2019). [g]Data on unpaid parking tickets accrued by diplomats in New York City are from the Fisman and Miguel (2007) study. [h]Corruption Perceptions Index (CPI) is a measure of the descriptive corruption norm from Transparency International's 2015 report. [i]The percentage of dropped wallets with money returned was taken from Figure 1 in the study by Cohn, Maréchal, Tannenbaum, and Zünd (2019). [j]Cultural distance was calculated following Kogut and Singh (1988) on the original four Hofstede dimensions (Power Distance, Individualism, Masculinity, and Uncertainty Avoidance; labeled "original") and on all six dimensions (labeled "all"). [k]Geographic Distance data were taken from the Centre d'Études Prospectives et d'Informations Internationales (CEPII) GeoDist database (Mayer & Zignago, 2012). Higher scores indicate a larger distance. [l]Linguistic-distance data were taken from the CCEPII Language database (Melitz & Toubal, 2014). Higher scores indicate greater difference in language. [m]These scores are based on genetic data from the Pemberton, DeGiorgio, and Rosenberg (2013) study, matched to country by Spolaore and Wacziarg (2018). Higher scores indicate a larger genetic distance.

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

freedom. The American scale has a smaller correlation with Long-Term Orientation, for which societies with more focus on the future are more culturally similar to the United States. The Chinese scale correlates with only 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 (2011) 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 (2015) measure is strongly correlated with the American scale of cultural distance, with tighter societies being 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 Chinese 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 (2006) 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 Chinese scale are more unreliable, although 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 that the United States is an outlier on these dimensions. None of the personality factors reliably correlate with the Chinese 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 Chinese 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 Chinese scale correlates with obedience and blood donations but not with the other psychological outcomes.

All distance measures show weak to moderate correlations with the American scale. With the exception of the traditional Kogut-Singh (1988) scale and the language-distance scale, most measures also correlate with the Chinese scale. The traditional Kogut-Singh scale derived from the original four Hofstede dimensions does not correlate with the Chinese scale, but the addition of the two variables that correlate with the Chinese scale increases the correlation. These correlations are reassuring in showing that the CF_{ST} approach to measuring cultural distance correlates in the right direction, with fewer high-resolution measures of cultural distance. The relationship between linguistic distance and the American scale was inconsistent, although 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 a proxy for cultural distance (Gorodnichenko & Roland, 2017) but is consistent with that found by other researchers (Giuliano, Spilimbergo, & Tonon, 2013), who argued that genetic distance captures geography but not culture.

Discussion

The psychological sciences face multiple crises, one of which is their overreliance on samples from 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 life span. We present a cultural-distance metric based on the F_{ST} technique from population genetics, applied to the WVS, a large survey of cultural values. 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 proven useful in answering questions in anthropology and economics (Bell et al., 2009; Desmet et al., 2017; Spolaore & Wacziarg, 2018). We anticipate that it will be equally useful to psychologists in addressing the WEIRD people problem.

To this end, we used the CF_{ST} to develop an American scale of cultural distance from the United States and a Chinese scale of cultural distance from China. CF_{ST} values can range from 0 to 1, yet all values on both scales are less than .3. This is consistent with findings of 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 Chinese 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, although 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 were dominated by Chinese psychologists, we would see a different set of psychological outcomes covered in textbooks, and these psychologies would correlate with a Chinese scale. Resolving which of these explanations is correct will require greater diversity in both researchers and samples.

Reassuringly, both the American scale and the Chinese scale correlate with proxies of cultural distance, suggesting that the scales are capturing some true cultural distance. In contrast, the original Kogut-Singh

(1988) distance is predictive only of the American scale. The addition of new dimensions with strong correlations to the Chinese scale improves the correlation with both the WEIRD and Chinese scales. We argue that by capturing only mean differences and ignoring differences in distributions, this alternative approach is systematically misleading.

The CF_{ST} (which has been made available through R code and at www.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 between any other grouping. We used this flexibility to compare the cultural differences between regions within the four largest populations—China, India, the United States, and the European Union. These analyses reveal that the cultural differences between regions of the overscrutinized United States are considerably smaller than those of the European Union, China, or India.

Researchers may disagree with our various decisions in constructing these scales. For example, we conservatively removed 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 the Answers as Alleles section), 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 that they believe are more defensible as culturally transmissible (see included-variables.csv in the Supplemental Material) or even argue that a different statistic is more appropriate (see Section 2 in the Supplemental Material). And indeed, biases may exist in the questions chosen for the WVS itself. We hope that the code we have provided 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 people 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—Japan, Hong Kong, and China (marked with asterisks)—are not 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 among 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 not only for 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 Supplemental Material, we offer some suggestions for developing these dimensions.

Relatively little attention has been paid to the Middle East and Africa by both the WVS (see gray regions in Fig. 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 that 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 have not 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.

Transparency

Action Editor: Ayse K. Uskul
Editor: D. Stephen Lindsay
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 the fixation index. M. Muthukrishna, C. M. Curtin, A. Gedranovich, J. McInerney, and B. Thue wrote code to analyze and graph the data. M. Muthukrishna drafted the manuscript. All the authors provided critical revisions and approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Funding for this study was provided by The London School of Economics and Political Science under the Suntory and

Toyota International Centres for Economics and Related Disciplines (STICERD) and Research and Innovation programs.

Open Practices

Data for this study were obtained from sources identified in the text. R code can be found in the Supplemental Material available online. The design and analysis plans for this study were not preregistered.

ORCID iD

Michael Muthukrishna https://orcid.org/0000-0002-7079-5166

Acknowledgments

We thank all of the individuals who provided comments about the draft version of this article on Twitter, various blogs, and other social media, and in particular, the detailed comments provided by R. McElreath, B. Purzycki, and other members of the Max Planck Institute for Evolutionary Anthropology. M. Muthukrishna also thanks Tyler Lieberthal for creating Figure 2.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620916782

Note

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).

References

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, *332*, 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, USA*, *106*, 17671–17674.

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

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton, NJ: 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*, 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; pp. 749–769). Hoboken, NJ: John Wiley and Sons. doi:10.1002/9781119125563.evpsych230

Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, *365*, 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*, 91–109.

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

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

Gelfand, M. (2019). *Rule makers, rule breakers: Tight and loose cultures and the secret signals that direct our lives*. New York, NY: 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*, 1100–1104.

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

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

Harrington, J. R., & Gelfand, M. J. (2014). Tightness–looseness across the 50 United States. *Proceedings of the National Academy of Sciences, USA*, *111*, 7990–7995.

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

Heine, S. J. (2015). *Cultural psychology* (3rd ed.). New York, NY: W. W. Norton.

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*, 903–918.

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 & Brain Sciences*, *33*, 61–83.

Hofstede, G. H. (n.d.). *The 6-D model of national culture*. Retrieved from https://geerthofstede.com/culture-geert-hofstede-gert-jan-hofstede/6d-model-of-national-culture/

Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Thousand Oaks, CA: SAGE.

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, USA*, *115*, 11428–11434.

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., et al. (Eds.). (2014). *World Values Survey: All Rounds - Country-Pooled Datafile 1981-2014*.

Retrieved from http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp

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

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., . . . Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*, 443–490.

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, *19*, 411–432.

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

Mayer, T., & Zignago, S. (2012). Notes on CEPII's distances measures: The GeoDist database (CEPII Working Paper No. 2011-25). SSRN. doi:10.2139/ssrn.1994531

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*, 407–425.

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

Muthukrishna, M., Francois, P., Pourahmadi, S., & Henrich, J. (2017). Corrupting cooperation and how anti-corruption strategies may backfire. *Nature Human Behaviour*, *1*, Article 0138. doi:10.1038/s41562-017-0138

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*, 221–229.

Muthukrishna, M., Morgan, T. J. H., & Henrich, J. (2016). The when and who of social learning and conformist transmission. *Evolution & Human Behavior*, *37*, 10–20.

Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*. Boulder, CO: 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*, 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, USA*, *115*, 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, USA*, *102*, 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*, 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), Article eaau5141. doi:10.1126/science.aau5141

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

Smaldino, P. E., Lukaszewski, A., von Rueden, C., & Gurven, M. (2019). Niche diversity can explain cross-cultural differences in personality structure. *Nature Human Behaviour*, *3*, 1276–1283.

Spolaore, E., & Wacziarg, R. (2018). Ancestry and development: New evidence. *Journal of Applied Econometrics*, *33*, 748–762.

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*, 603–608.

Transparency International. (2016). *Corruption Perceptions Index 2015*. Retrieved from https://www.transparency.org/cpi2015

Twain, M. (1989). *Following the equator: A journey around the world*. Hartford, CN: Dover. (Original work published 1897)

Uz, I. (2015). The index of cultural tightness and looseness among 68 countries. *Journal of Cross-Cultural Psychology*, *46*, 319–335.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). New York, NY: Springer.

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