1 The non-independence of nations and why it matters

2 Scott Claessens[1] & Quentin D. Atkinson[1,2]

3 [1] School of Psychology, University of Auckland, Auckland, New Zealand

4 [2] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

5 Author Note

Abstract

Cross-national analyses test hypotheses about the drivers of global variation in national outcomes. However, since nations are connected in various ways, such as via spatial proximity and shared cultural ancestry, cross-national analyses often violate assumptions of non-independence, inflating false positive rates. Here, we show that, despite being recognised as an important statistical pitfall for over 200 years, cross-national research in economics and psychology still does not sufficiently account for non-independence. In a review of the 100 highest-cited cross-national studies of economic development and values, we find that controls for non-independence are rare. When studies do include controls for non-independence, our simulations suggest that commonly used methods continue to produce false positives. In reanalyses of twelve cross-national relationships, we show that half are no longer significant after controlling for non-independence using global proximity matrices. We urge social scientists to sufficiently control for non-independence in cross-national research.

Keywords: spatial non-independence; cultural non-independence; cross-national analyses; simulations; replications; reanalyses

Word count: 5459 words

30   The non-independence of nations and why it matters

31   Nations are perhaps the single most important population unit structuring modern

32   human life. The nation in which someone is born has a large effect on what they can expect

33   out of life, including their income level[1], life expectancy[2], mental health[3], subjective

34   well-being[4], and educational attainment[5]. Nations are also among the most important

35   determinants of human cultural variation, with greater cultural similarity within than

36   outside national borders[6].

37   Given the importance of nations for structuring human behaviour, there is justifiably

38   huge interest in statistical analyses that attempt to predict variation in national outcomes of

39   all kinds. At the time of writing, a search in Web of Science for the terms "cross-national",

40   "cross-cultural", or "cross-country" in titles and abstracts returned 80,000 unique hits. The

41   standard practice for cross-national analyses is to conduct bivariate correlations or multiple

42   regressions with individual data points representing different nations. Such analyses widen

43   the scope of social science beyond Western populations[7,8] and have been used to study,

44   among other topics, the causes of variation in the economic wealth of nations[9–12], global

45   patterings of cultural norms and values[13–16], and the universality and diversity of human

46   behaviour and psychology around the world[17–20].

47   However, cross-national analyses are complicated by the fact that nations are not

48   statistically independent data points. Unlike independent random samples from a

49   population, nations are related to one another in a number of ways. First, nations that are

50   closer to one another tend to be more similar than distant nations. This phenomenon is

51   known as spatial non-independence[21], and it occurs because nations in close spatial

52   proximity share characteristics due to local cultural diffusion of ideas[22] and regional

53   variation in climate and environment[21]. For example, the neighbouring African nations

54   Zambia and Tanzania have similar levels of terrain ruggedness, which has been used to

55   partially explain their similar levels of economic development[23]. This pattern conforms to

Tobler's first law of geography: "everything is related to everything else, but near things are more related than distant things"[24] (p. 236).

Second, nations with shared cultural ancestry tend to be more similar than culturally unrelated nations. This is known as cultural phylogenetic non-independence[25–27], and occurs because related nations share cultural traits inherited via descent from a common ancestor. Shared cultural ancestry can result in a form of pseudoreplication, whereby multiple instances of the same trait across nations are merely duplicates of the ancestral original. For example, despite being separated by over 1,500 kilometres of ocean, the related island nations Tonga and Tuvalu share similar languages and customs due to cultural inheritance from a common Polynesian population dating back more than 1,000 years. More recently, shared ancestry explains a myriad of cultural similarities between colonial settlements and their colonisers (e.g. Argentina and Spain).

Spatial and cultural phylogenetic non-independence between nations make cross-national inference challenging. A fundamental assumption of regression analysis is that model residuals should be independently and identically distributed[28]. But without accounting for spatial or cultural non-independence between nations, model residuals can show structure that remains unaccounted for, violating this assumption. Treating nations as independent can thus inflate false positive rates[29], producing spurious "direct" relationships between variables that in fact only indirectly covary due to spatial or cultural connections[30] (see Supplementary Figure S1 for an example causal model).

Non-independence between data points is widely acknowledged in fields that routinely deal with spatially or culturally structured data. In ecology and sociology, studies with regional-level data use a variety of autoregressive models designed to account for spatial patternings[31,32]. In anthropology, researchers have recognised cultural non-independence as an important statistical pitfall for over 200 years, with issues of cultural pseudoreplication being identified in early comparative studies of marriage practices across societies[25]. More

82 recently, the Standard Cross-Cultural Sample of 186 cultures was compiled to minimise the

83 confounding effects of non-independence in comparisons of traditional societies[33], though

84 spatial and cultural dependencies are difficult to remove entirely[34,35]. Anthropologists also

85 borrow phylogenetic comparative methods from evolutionary biology, such as phylogenetic

86 least squares regression[36], when comparing traditional societies, treating culturally related

87 societies in the same way as biologists treat genetically related species (e.g.[37,38]).

88 At the national level, recent reanalyses have revealed that several cross-national

89 relationships reported in economics and psychology do not hold when controlling for

90 non-independence between nations. One study replicated 25 analyses of "persistence" in

91 economics, in which modern national outcomes are regressed against historical characteristics

92 of those nations, and found that over half of the relationships were attenuated when

93 controlling for spatial non-independence[39]. Another replication study found that many of the

94 widely publicised relationships between national-level pathogen prevalence and political

95 institutions and attitudes fail to hold when controlling for various kinds of

96 non-independence[40]. These reanalyses, and others[41–43], raise the question: how widespread a

97 concern is non-independence in studies of national-level outcomes?

98 To address this question, we consider national-level variables of general interest across

99 the social sciences: economic development and cultural values. These variables are frequently

100 included as both outcomes and predictors in cross-national studies in economics and

101 psychology[9–16]. First, we show that economic development and cultural values are spatially

102 and culturally non-independent across nations, emphasising the need to control for

103 non-independence. Second, we systematically review the 100 highest-cited cross-national

104 studies of economic development and cultural values and estimate the proportion of

105 cross-national analyses within these articles that account for non-independence between

106 nations. Third, we run simulations to determine whether common methods of dealing with

107 non-independence in the literature sufficiently reduce false positive rates. Fourth, we

108 reanalyse twelve previous cross-national analyses of economic development and cultural

109 values from our systematic review, incorporating global geographic and linguistic proximity

110 matrices to correctly control for spatial and cultural non-independence.

<h1 style="text-align:center">Results</h1>

111

National-level economic development and cultural values are spatially and culturally non-independent

114   In order to motivate our research question, it is important to first show that economic

115 development and cultural values are indeed spatially and culturally non-independent around

116 the world. To this end, we used Bayesian multilevel models to simultaneously estimate

117 geographic and cultural phylogenetic signal for the Human Development Index[44] and two

118 primary dimensions of cultural values from the World Values Survey, traditional vs. secular

119 values and survival vs. self-expression values[16]. For all three of these variables, we found that

120 a substantial proportion of national-level variation was explained by spatial proximity and

121 shared cultural ancestry between nations (Figure 1). In our models, shared cultural ancestry

122 explained over half of the national-level variation in economic development and cultural

123 values, with spatial proximity explaining most of the remaining national-level variance.

124 Bayes Factors indicated strong evidence that these estimates of geographic and cultural

125 phylogenetic signal differed from zero, aside from the geographic signal for survival values,

126 for which the evidence was equivocal (see Supplementary Results). These findings emphasise

127 the need to account for spatial and cultural phylogenetic non-independence in cross-national

128 analyses of economic development and cultural values.

Previous cross-national analyses have not sufficiently accounted for non-independence

131   Given that economic development and cultural values show evidence of geographic and

132 cultural phylogenetic signal, have cross-national analyses sufficiently accounted for this
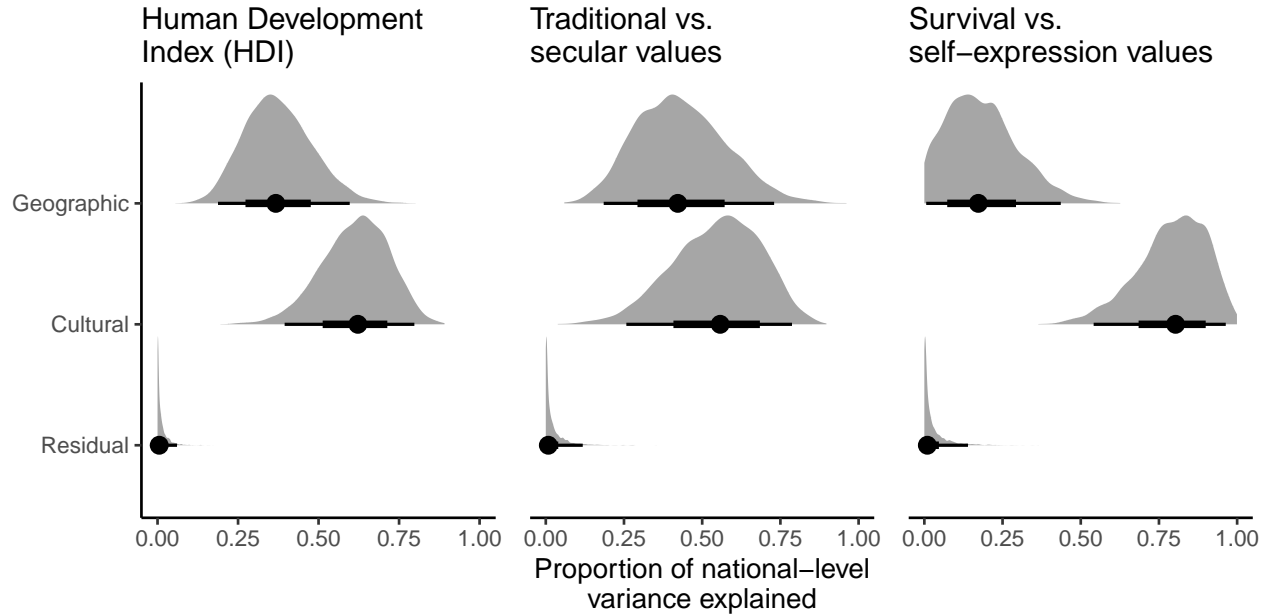
Figure 1. *Posterior estimates of geographic and cultural phylogenetic signal for the Human Development Index and Inglehart's traditional and survival values.* Geographic and cultural phylogenetic signal are operationalised as the proportion of national-level variance explained by geographic and linguistic proximity matrices. Grey ridges are full posterior distributions, points are posterior median values, and lines are 50% and 95% credible intervals.

non-independence? To assess this, we systematically searched the published literature for articles that combined the search terms "economic development" or "values" with the search terms "cross-national", "cross-cultural", or "cross-country". We removed articles that did not report original research, were not relevant to economic development or cultural values, or did not report at least one cross-national analysis. We then retained the 100 articles (50 for economic development, 50 for cultural values) with the highest annual rate of citations. For each of these highly-cited articles, we exhaustively recorded every cross-national analysis reported in the main text (n = 4,308), identifying in each case whether or not the analysis attempted to control for spatial, cultural, or any other form of non-independence between nations (see Methods for detailed search criteria and coding decisions).

The results of our systematic review show that most published articles and

cross-national analyses make no attempt to account for statistical non-independence. Figure 2a plots the proportion of articles that contain at least one cross-national analysis accounting for non-independence. We find that 42% of economic development articles contain at least one attempt to control for non-independence (95% bootstrap confidence interval [0.30 0.54]), while this proportion decreases to only 8% for cultural values articles (95% bCI [0.02 0.16]). Both kinds of article are most likely to use regional fixed effects (e.g. continent fixed effects) to account for non-independence, but some articles also include controls for spatial distance (e.g. latitude) and shared cultural history (e.g. colony status).

Focusing on the full sample of 4,308 analyses, we find that the proportion of individual cross-national analyses accounting for non-independence is even lower (Figure 2b). Across 2,487 cross-national analyses from studies of economic development, only 5% are estimated to control for non-independence (95% credible interval [0.02 0.15]). Similarly, across 1,821 cross-national analyses from studies of cultural values, only 1% are estimated to control for non-independence (95% CI [0.00 0.02]).

Since our systematic review goes back as far as 1993 (Figure 2c), it is possible that our estimates are being biased by earlier studies, and that controls for non-independence have increased over time with methodological advancements and greater awareness of the issue. To test this possibility, we fitted a time trend to the full sample of analyses with a Bayesian multilevel spline model. However, we found that, for both studies of economic development and cultural values, the estimated probability of controlling for non-independence has remained low since 1993 (Figure 2d).

Common methods of controlling for non-independence produce inflated false positive rates

Our systematic review revealed that most cross-national analyses in the literature do not control for spatial or cultural phylogenetic non-independence. When they do, they tend
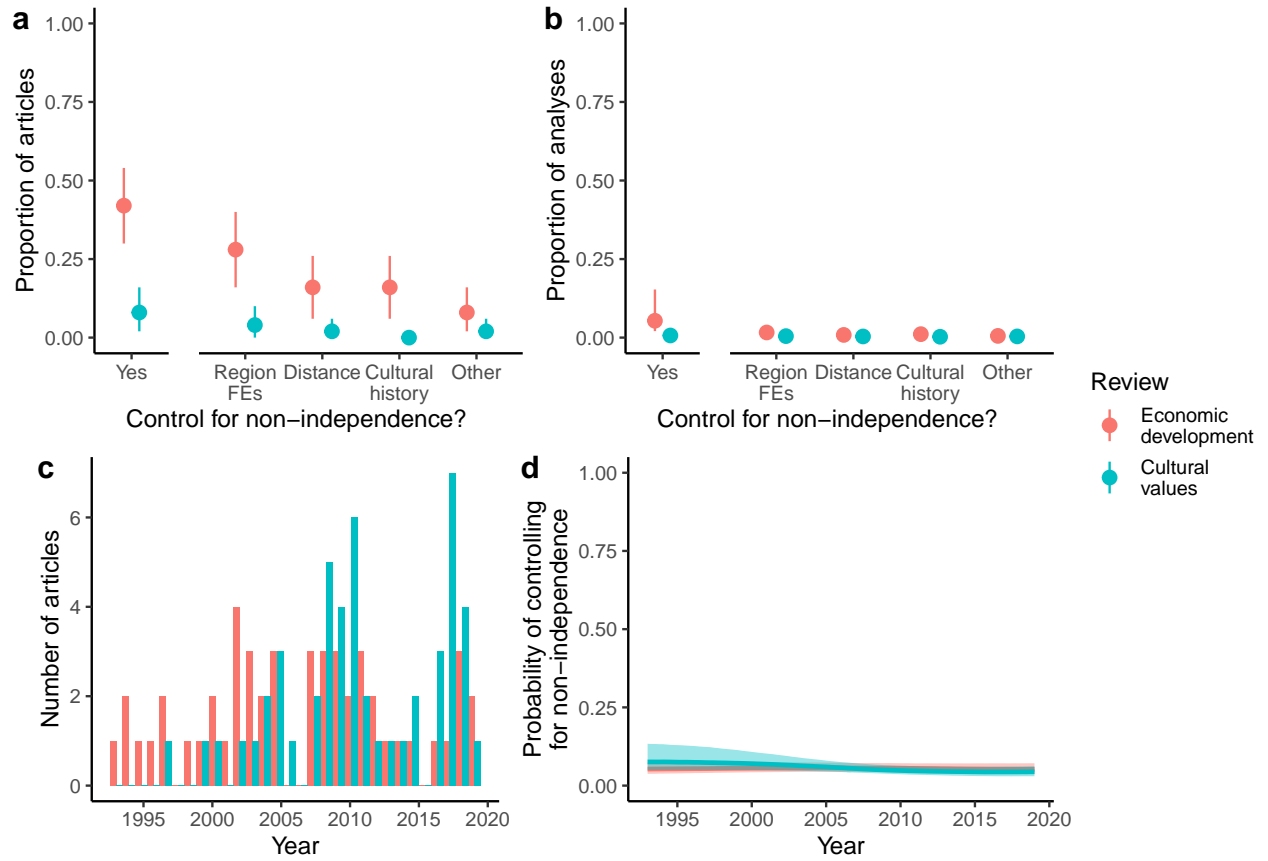
Figure 2. *Results from systematic review of 100 highly-cited cross-national studies of economic development (red) and cultural values (blue).* (a) Proportion of articles containing at least one analysis accounting for non-independence, overall and split by common methods of controlling for non-independence. Point ranges represent proportions and 95% bootstrap confidence intervals. (b) Adjusted proportion of individual analyses accounting for non-independence, overall and split by method. Point ranges are posterior medians and 95% credible intervals from Bayesian multilevel logistic regressions. (c) Histogram of publication years for studies of economic development and cultural values. (d) Estimated trend over time for the probability of controlling for non-independence. Lines and shaded areas are posterior median regression lines and 50% credible intervals from a Bayesian multilevel spline model. Region FEs = region fixed effects.

169 to include controls like latitude and regional fixed effects. Do these methods sufficiently

170 account for statistical non-independence?

171   To compare the efficacy of different methods in the literature, we conducted a

172 simulation study. We simulated national-level datasets (n = 236 nations) with varying

173 degrees of spatial or cultural phylogenetic autocorrelation (i.e. non-independence) for

174 outcome and predictor variables, but with no direct causal relationship between the variables.

175 We then fitted naive regressions without controls to these datasets, as well as regression

176 models with controls for latitude, longitude, and continent fixed effects. Despite not being

177 identified in our systematic review, we also included fixed effects for the language families of

178 the majority-spoken languages in each country, as this control is often used to account for

179 cultural phylogenetic non-independence (e.g.[45]). In addition, we included Conley standard

180 errors, a widely used standard error correction that purportedly accounts for spatial

181 non-independence[46,47]. Finally, we included Bayesian multilevel regressions that explicitly

182 model spatial and/or cultural phylogenetic non-independence by allowing nation random

183 intercepts to covary according to global geographic and/or linguistic proximity matrices (see

184 Supplementary Methods). To model spatial non-independence we included a Gaussian

185 process[48,49] over latitude and longitude values, and to model cultural phylogenetic

186 non-independence we assumed that nation random intercepts were correlated in proportion

187 to their linguistic proximity[50]. Across all model types, false positive rates were measured as

188 the proportion of models that estimated a slope with a 95% confidence / credible interval

189 excluding zero (i.e. falsely infering a relationship when none is present).

190   Figures 3 and 4 plot the estimated false positive rates from our simulation study, split

191 by different methods and different degrees of spatial or cultural phylogenetic autocorrelation

192 (see Supplementary Figures S2 and S3 for full distributions of effect sizes under strong

193 autocorrelation). For reference, "weak" autocorrelation in our simulation is comparable to

194 the geographic signal for survival values in Figure 1, while "moderate" and "strong" levels of

autocorrelation are comparable to the cultural phylogenetic signal for traditional and survival values, respectively.

Our simulation study revealed that with at least moderate degrees of spatial or cultural phylogenetic autocorrelation for both outcome and predictor variables, naive regression models produce false positive rates above chance levels. This false positive rate increases as the degree of autocorrelation increases. With strong spatial autocorrelation for both outcomes and predictors, false positive rates reach as high as 71%. We find a slightly lower false positive rate under strong cultural phylogenetic autocorrelation, though this false positive rate is still greater than expected by chance (38%).

Common methods in the literature do not reduce these high false positive rates. With strong spatial autocorrelation for both outcome and predictor variables, false positive rates remain above 50% when controlling for latitude, longitude, and language family fixed effects (Figure 3). Applying Conley standard errors also does not reduce false positive rates below 50% under strong spatial autocorrelation, regardless of the distance cutoff. Continent fixed effects are more effective than other frequentist methods, though they continue to produce a false positive rate of 28% under strong spatial autocorrelation. By contrast, Bayesian spatial Gaussian process regression with longitude and latitude values outperforms all other methods. This approach eliminates false positives under moderate spatial autocorrelation, such that the false positive rate is no different from chance, and reduces the false positive rate under strong spatial autocorrelation to 17%. Bayesian models that additionally account for linguistic proximity between nations perform equally well, though models with only linguistic covariance continue to produce false positives.

In our simulation of cultural phylogenetic non-independence, we find that most frequentist methods do not reduce false positive rates (Figure 4). Controls for latitude and longitude, continent fixed effects, and Conley standard errors do little to change false positive rates. Language family fixed effects are slightly more effective than other frequentist
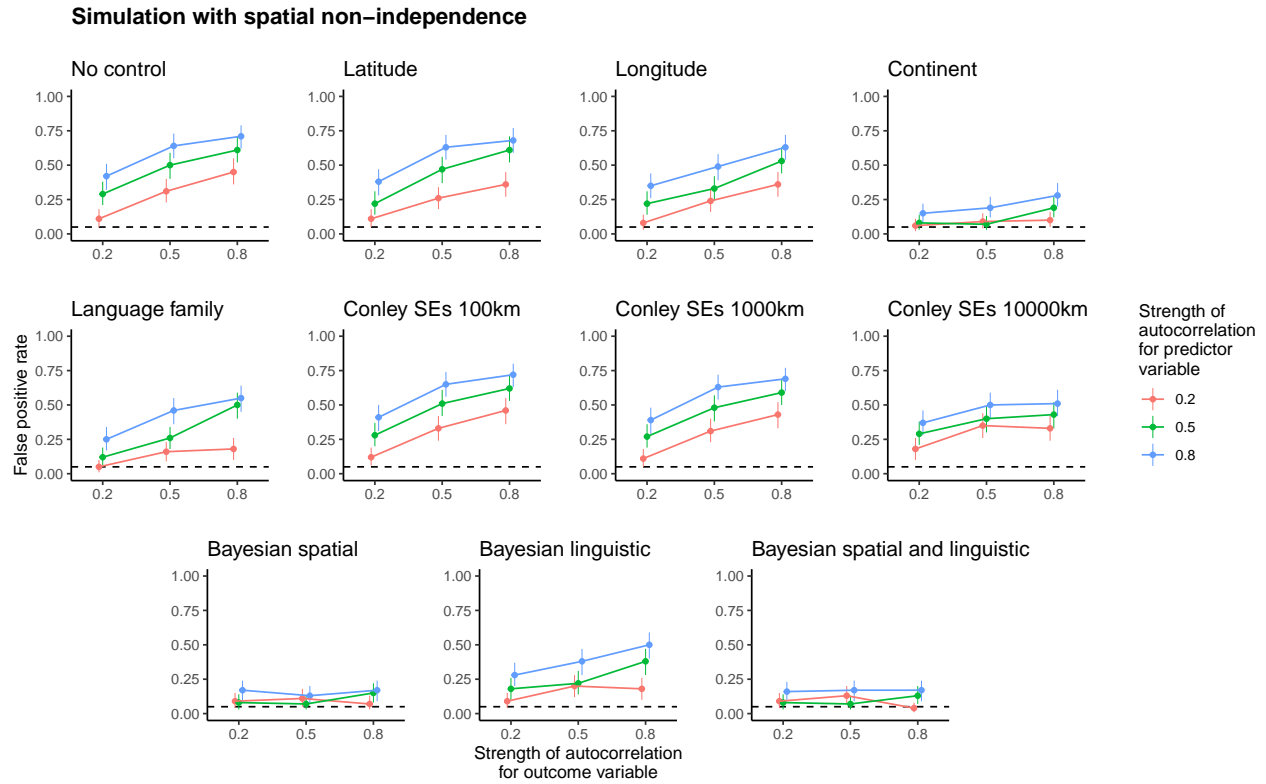
Figure 3. *False positive rates for different methods of controlling for spatial non-independence in our simulation study.* For simulated outcome and predictor variables, we systematically varied the strength of spatial autocorrelation, from weak (0.2) to moderate (0.5) to strong (0.8). We simulated 100 datasets per parameter combination and fitted different models to each dataset. False positive rates were operationalised as the proportion of models that estimated a slope with a 95% confidence / credible interval excluding zero. Point ranges represent proportions and 95% bootstrap confidence intervals, and dashed lines indicate the 5% false positive rate that is expected due to chance. SEs = standard errors.

Figure 4. *False positive rates for different methods of controlling for cultural phylogenetic non-independence in our simulation study.* For simulated outcome and predictor variables, we systematically varied the strength of cultural phylogenetic autocorrelation, from weak (0.2) to moderate (0.5) to strong (0.8). We simulated 100 datasets per parameter combination and fitted different models to each dataset. False positive rates were operationalised as the proportion of models that estimated a slope with a 95% confidence / credible interval excluding zero. Point ranges represent proportions and 95% bootstrap confidence intervals, and dashed lines indicate the 5% false positive rate that is expected due to chance. SEs = standard errors.
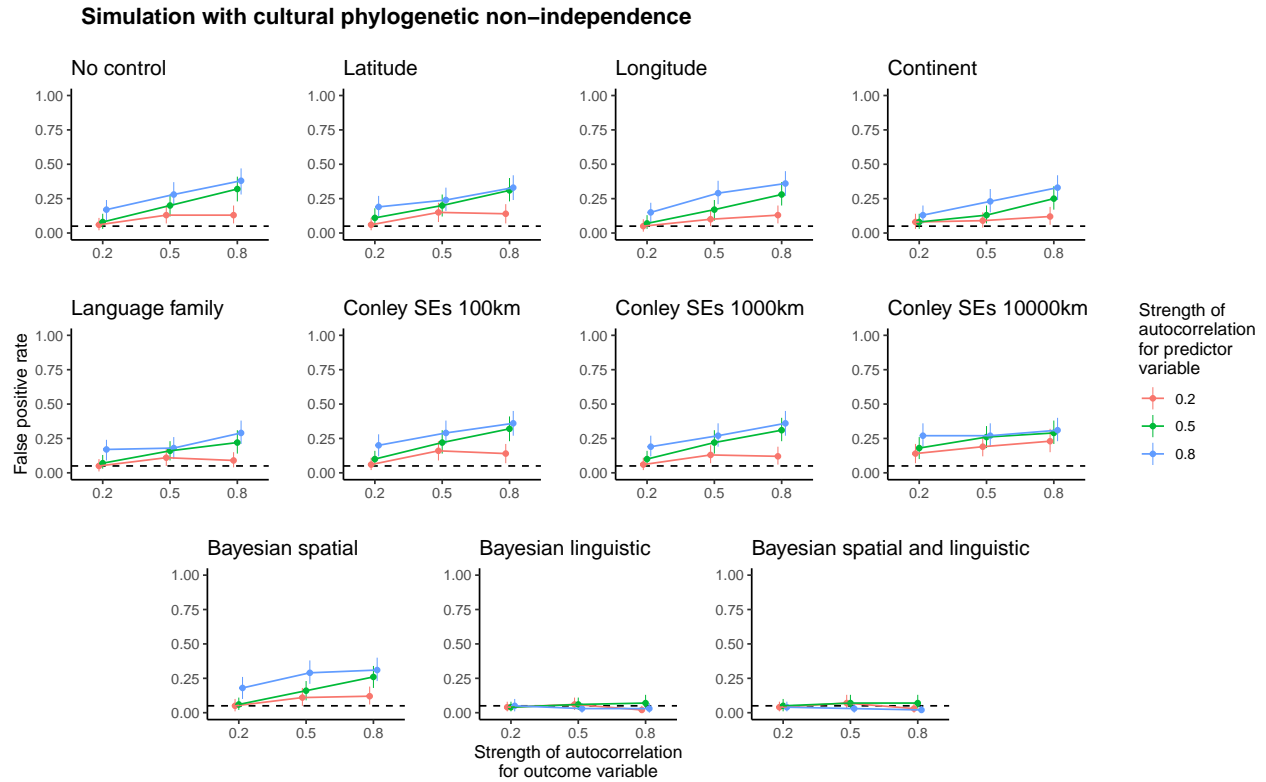
methods, though they continue to produce a false positive rate of 29% under strong cultural phylogenetic autocorrelation. By contrast, Bayesian models with random effects covarying according to linguistic proximity completely eliminate false positives across all degrees of cultural phylogenetic autocorrelation. Bayesian models that additionally account for geographic proximity between nations perform equally well, though models with only a spatial Gaussian process continue to produce false positives.

Key findings in the literature are not robust to reanalysis with more rigorous methods

Our systematic review and simulation study have shown that controls for non-independence are rare in cross-national studies of economic development and cultural values, and when studies do attempt to control for non-independence, the methods typically used are likely to continue to produce false positives. This raises the worrying possibility that the cross-national literature in economics and psychology is populated with spurious relationships.

To determine how widespread this issue of spurious cross-national relationships might be, we reanalysed a subset of twelve previous cross-national analyses from our systematic review, sufficiently controlling for spatial and cultural phylogenetic non-independence using global geographic and linguistic proximity matrices. We subsampled six analyses from our economic development review[51–56] and six from our cultural values review[13,14,16,57–59]. Our choice of analyses was constrained by data availability and whether we were able to initially replicate the original finding. We pre-registered our subsample of analyses before running any control models (https://osf.io/uywx8/). We controlled for non-independence by including (1) a Gaussian process allowing nation random intercepts to covary according to a geographic proximity matrix, and/or (2) nation random intercepts that covaried according to a linguistic proximity matrix (see Supplementary Methods for full models).

246   Figure 5 visualises the results of our reanalysis. Cross-national correlation effect sizes

247 tended to reduce when controlling for statistical non-independence between nations,

248 sometimes by as much as half of the original effect size. Overall, after controlling for

249 non-independence, six out of twelve cross-national associations had 95% credible intervals

250 that included zero. For the economic development analyses, four out of six cross-national

251 relationships had 95% credible intervals including zero when controlling for spatial

252 non-independence. For the cultural values analyses, two out of six cross-national

253 relationships had 95% credible intervals including zero when controlling for cultural

254 phylogenetic non-independence. Supplementary Figure S4 shows these cross-national

255 correlations plotted against the raw data.

256   To understand why some cross-national correlations were attenuated by controls for

257 non-independence while others were robust, we further explored our fitted models for

258 evidence of spatial and cultural autocorrelation. For each outcome variable, our Gaussian

259 process models provided varying estimates of how quickly spatial autocorrelation declined

260 with distance (Supplementary Figure S5). For example, in Skidmore and Toya[56] gross

261 domestic product growth was only moderately spatially autocorrelated at 1,000 km distance

262 (posterior median spatial autocorrelation at 1,000 km = 0.47, 95% CI [0.06 0.97]), whereas in

263 Inglehart and Baker[16] traditional values were strongly spatially autocorrelated at the same

264 distance (posterior median spatial autocorrelation at 1,000 km = 0.96, 95% CI [0.78 0.99]).

265 We also found varying estimates of cultural phylogenetic signal (Supplementary Figure S6),

266 with some outcome variables expressing low signal (e.g. tightness[14]; posterior median = 0.11,

267 95% CI [0.00 0.82]) and others expressing high signal (e.g. female labour force

268 participation[58]; posterior median = 0.91, 95% CI [0.67 0.99]). Across all analyses, we found

269 that stronger estimates of spatial autocorrelation or cultural phylogenetic signal resulted in a

270 more pronounced reduction in the effect size when controlling for non-independence between
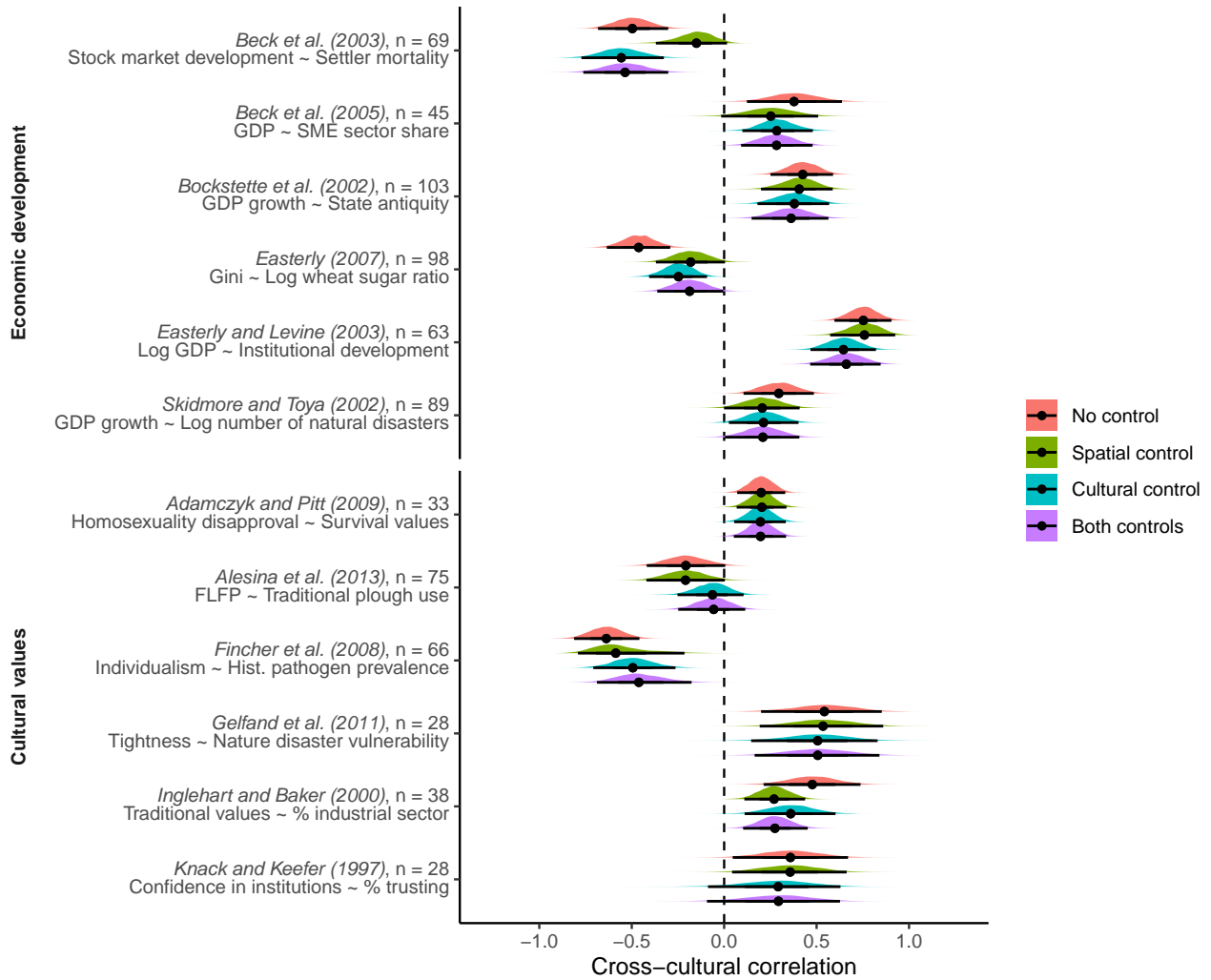
271 nations (Figure 6).

Figure 5. *Posterior correlations from our reanalysis of twelve previous cross-national analyses.* For each previous cross-national relationship, we plot the posterior slopes from a naive regression (red), a regression controlling for spatial non-independence (green), a regression controlling for cultural phylogenetic non-independence (blue), and a regression controlling for both spatial and cultural phylogenetic non-independence simultaneously (purple). All outcome and predictor variables are standardised. Most analyses are simple bivariate cross-national correlations, but Gelfand et al. (2011) is a partial correlation controlling for log gross national income and Adamczyk and Pitt (2009) is a multilevel model including several covariates. Point ranges represent posterior medians and 95% credible intervals. GDP = gross domestic product. FLFP = female labour force participation.
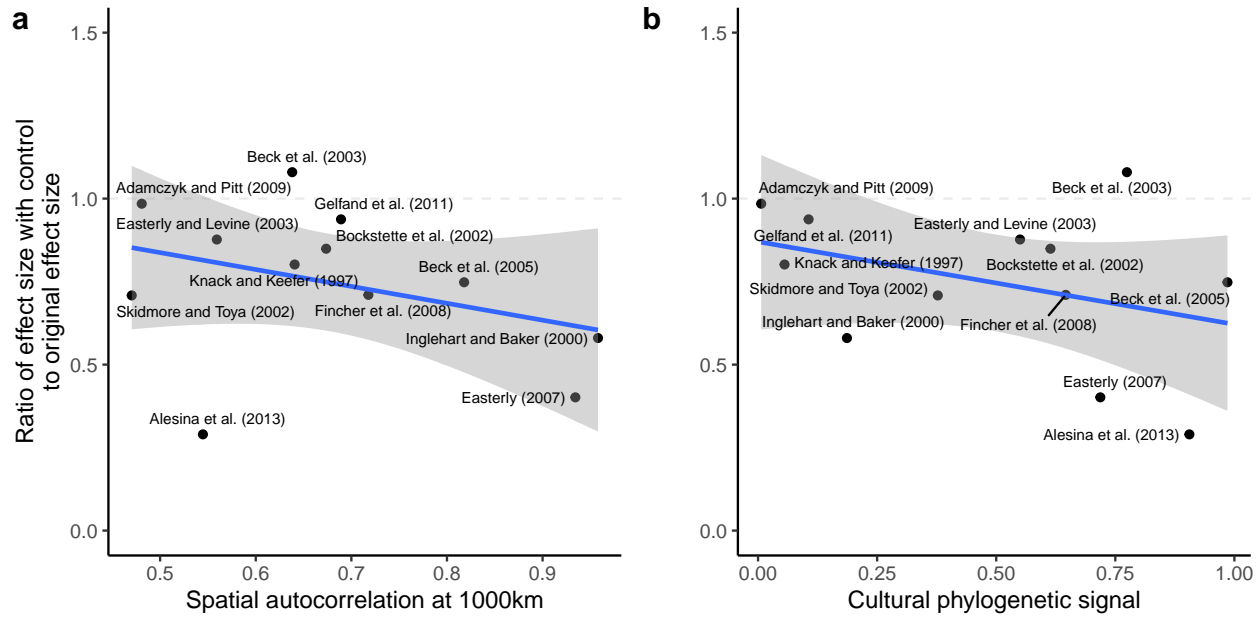
Figure 6. *The estimated degree of spatial and cultural phylogenetic non-independence predicts reductions in effect size in our reanalysis.* (a) Higher estimated degrees of spatial autocorrelation at 1,000 km distance predict more pronounced reductions in effect sizes when controlling for non-independence. (b) Higher estimated levels of cultural phylogenetic signal predict more pronounced reductions in effect sizes when controlling for non-independence. In both panels, the y-axis represents the ratio of the effect size when controlling for spatial and cultural non-independence to the original effect size (from naive regression model), and the x-axis represents posterior median model estimates. Regression lines are plotted with 95% confidence intervals.

Discussion

In a systematic literature review and simulation, we found that cross-national studies in economics and psychology rarely account for non-independence between nations, and, when they do, the methods they use continue to produce false positives. In a reanalysis of twelve cross-national correlations, we further showed that neglecting to account for non-independence has resulted in spurious relationships in the published literature, with half of the correlations failing to replicate when controlling for spatial or cultural non-independence with more rigorous methods. These findings suggest that cross-national

280 analyses in economics and psychology should be interpreted with caution until

281 non-independence is sufficiently accounted for.

282   Our initial analyses add to and clarify existing evidence regarding the

283 non-independence of economic and cultural variation among nations. One previous study

284 suggested that geographic proximity is more important than deep cultural ancestry in

285 explaining the distribution of human development across Eurasian nations, though the

286 authors noted that their small sample of 44 nations and regional focus limited their

287 statistical power[60]. Our global sample of nations revealed strong cultural phylogenetic signal,

288 as well as geographic signal, for the Human Development Index. Another previous study

289 found that similarities in the cultural values of nations are predicted by linguistic, but not

290 geographic, distances between those nations[6]. We find this same result for survival

291 vs. self-expression values, but for traditional vs. secular values we find that both linguistic

292 and geographic proximity are important independent predictors of global variation. These

293 findings emphasise the need to account for both spatial and cultural phylogenetic

294 non-independence in cross-national studies of economic development and cultural values.

295   Crucially, our systematic literature review and simulation study revealed that the most

296 commonly used controls for non-independence do not sufficiently deal with the issue. In our

297 simulations, controlling for either latitude or longitude did not reduce false positive rates.

298 This result calls into question controls like distance to the equator to account for

299 non-independence in cross-national regression models, though these controls may still be

300 suitable to account for regional or latitudinal variation in ecology, which we did not simulate.

301 High false positive rates persisted with Conley standard errors, which have previously been

302 critiqued for being overly sensitive to arbitrary distance cutoffs[39]. The simulation also

303 confirmed the assertion that fixed effects for spatial or cultural groupings (e.g. continent or

304 language family fixed effects) are insufficient because non-independence still remains within

305 groupings[40]. This logic further applies to analyses that control for non-independence by

306 separately analysing different regions (e.g.[61]). Across all model types in our simulation, the
307 only methods that sufficiently reduced the false positive rate were the Bayesian multilevel
308 regressions that explicitly modelled spatial and cultural phylogenetic autocorrelation, though
309 we did not include other possible controls for non-independence, such as conditional
310 autoregressive models[31] or generalised additive models[62].

311   Ours is not the first review to show that studies are misapplying statistical methods in
312 ways that inflate false positive rates. For example, other literature reviews have shown that
313 studies in the social sciences tend to use small samples of participants[63], treat ordinal data
314 as metric[64], incorrectly handle missing values[65], and ignore best practices in meta-analyses[66].
315 Why do cross-national studies also rarely account for non-independence? At the institutional
316 level, one possibility is that such practices are incentivised *because* they generate statistically
317 significant relationships, which increase the probability that a study is published[63]. At the
318 individual level, another possibility is that researchers outside of anthropology and ecology
319 are simply not aware of the problem, or believe that the problem does not apply to analyses
320 of nations. Even if researchers appreciate the problem, they might not know of suitable
321 controls or perceive the methods to be too complex.

322   These institutional- and individual-level barriers can be combatted. First,
323 cross-national replication studies like ours and others[39–43], combined with the methodological
324 reviews included in Registered Reports[67], might change incentive structures and encourage
325 researchers to analyse the world's nations with more rigorous methods. Second, since the
326 issue of non-independence is fundamentally an issue of causal inference (Supplementary
327 Figure S1), more explicit descriptions of causal models could promote controls for
328 non-independence in cross-national research. In our review, economists studying economic
329 development dealt with national-level non-independence more than psychologists studying
330 cultural values, likely because economics studies tend to be lengthy statistical exercises that
331 systematically incorporate or exclude numerous variables in an attempt to infer causation.

Third, the recent widespread accessibility of open source statistical software, such as the programming language Stan[68] and the R package *brms*[69], should promote the use of more rigorous methods to control for non-independence. Using *brms*, for example, Bayesian Gaussian process regression is straightforward to conduct, requiring only longitude and latitude values for nations.

Until such changes are implemented and sufficient controls for non-independence are the norm, existing cross-national correlations should be interpreted with caution. In our reanalyses, we found that four out of six cross-national correlations with economic development variables had 95% credible intervals that included zero when controlling for spatial non-independence. Three of these analyses were tests of "persistence" hypotheses, studying the effects of historical and environmental conditions — settler mortality[51], wheat-sugar suitability[55], and natural disaster frequency[56] — on modern developmental outcomes. A recent reanalysis has also called into question various studies of this ilk[39]. We also found that two out of six cross-national correlations with cultural values variables had 95% credible intervals that included zero when controlling for cultural phylogenetic non-independence.

We do not wish to dissuade researchers from conducting cross-national studies. On the contrary, such work promises to deepen understanding of our world, including the causes and consequences of economic development and cultural values. Moreover, cross-national studies allow social scientists to broaden their scope of study beyond Western populations[7], providing the representative samples necessary to test evolutionary and socio-ecological theories of human behaviour[8,70]. But in order to minimise spurious relationships in global datasets, we urge researchers to control for spatial and cultural phylogenetic non-independence when reporting cross-national correlations. Nations are not independent, and our statistical models must reflect this.

Methods

Geographic and cultural phylogenetic signal

To estimate the degree of spatial and cultural phylogenetic non-independence in economic development and cultural values, we calculated geographic and cultural phylogenetic signal for global measures of development and values. Our measure of economic development was the Human Development Index[44]. We retrieved a longitudinal dataset capturing human development for 189 nations since 1990 (n = 1,512; https://hdr.undp.org/en/content/download-data). Our measures of cultural values were traditional vs. secular values and survival vs. self-expression values from the World Values Survey[16]. We downloaded the full Integrated Values Survey, which included all waves from the World Values Survey and the European Values Survey, and computed the two dimensions of cultural values following procedures from previous research[16]. This longitudinal dataset captures values and attitudes for 116 nations since 1981 (n = 645,249; https://www.worldvaluessurvey.org/WVSEVStrend.jsp).

To calculate geographic and cultural phylogenetic signal, we created two proximity matrices for 269 of the world's nations: a geographic proximity matrix and a linguistic proximity matrix. Geographic proximity was converted from logged geodesic distances between nation capital cities. Linguistic proximity was calculated as the cultural proximity between all languages spoken within nations, weighted by speaker percentages (see Supplementary Methods). We included these matrices in Bayesian multilevel models, allowing nation random intercepts to covary according to both geographic and linguistic proximity simultaneously. These models were fitted with the R package *brms*[69] and converged normally ($\hat{R} < 1.1$). Estimates of geographic and cultural phylogenetic signal were computed as the proportion of national-level variance in these models explained by geographic and linguistic proximity matrices.

Systematic literature review

We exported two searches from Web of Science (https://www.webofknowledge.com/) on 27th September 2021, restricting our searches to articles published between 1900 and 2018. The first search was for the terms "economic development" AND ("cross-national" OR "cross-cultural" OR "cross-country"), which returned 965 articles. The second search was for the terms "values" AND ("cross-national" OR "cross-cultural" OR "cross-country"), which returned 6806 articles. Once exported, we ordered the articles by descending number of citations per year since initial publication, using citation counts reported by Web of Science.

We then systematically coded each article, in order, for inclusion in our review. Articles were only included if: (1) they were judged to be relevant to economic development or cultural values; (2) they were an original empirical research article; and (3) they contained at least one analysis with national-level outcome or predictor variables. We stopped when we had included 50 articles for the economic development review and 50 articles for the cultural values review.

Within each included article, we exhaustively coded every individual cross-national analysis reported in the main text. We coded mainly correlation or regression analyses, and explicitly excluded meta-analyses, factor analyses, measurement invariance analyses, multidimensional scaling analyses, hierarchical clustering analyses, multiverse analyses, and scale development / validation analyses. We also excluded analyses that compared only two, three, four, five, or six nations. For each included analysis, we recorded the year, outcome variable, all predictor variables, test statistic, p-value, number of nations, number of data points, model type, if the data were available, and whether and how the analysis attempted to control for non-independence.

We coded common attempts to control for non-independence between nations. These included: (1) any higher-level control variables for spatial regional groupings (e.g. continent

fixed effects); (2) any geographic distance control variables (e.g. distance between capital

cities, distance from equator, latitude); (3) any control variables capturing shared cultural

history (e.g. former colony, legal origin fixed effects, linguistic history, cultural influence);

and (4) any other control variables, tests, or approaches that were deemed as attempts to

control for non-independence (e.g. eigenvector filtering[71], controls for trade-weightings

between countries, cross-sectional dependence tests[72], separate analyses for subsets of

countries). These were coded by the first author.

Once we had compiled our review database, we calculated the proportion of articles

attempting to control for non-independence at least once. We also calculated the proportion

of articles employing the different types of control listed above at least once: regional fixed

effects, distance, shared cultural history, or other. For these proportions, we calculated 95%

boostrap confidence intervals with 1,000 bootstrap iterations.

For individual analyses, we dealt with the nested nature of the data (analyses nested

within articles) by fitting Bayesian multilevel logistic regression models with review type

(economic development vs. cultural values) as the sole fixed effect and random intercepts for

articles. We fitted these models separately for overall attempts to control for

non-independence and split by method type. We report the adjusted proportions with 95%

credible intervals. To test for a trend over time, we also fitted a Bayesian multilevel logistic

regression with a multigroup spline for year of publication and random intercepts for articles.

Bayesian models were fitted with the *brms* R package[69]. Our priors were informed by prior

predictive checks, and all models converged normally ($\hat{R} < 1.1$).

Simulations

We simulated data for 236 nations with varying degrees of spatial or cultural

phylogenetic signal for outcome y and predictor x using the following generative model:

$$y = \alpha_y + \epsilon_y$$

$$x = \alpha_x + \epsilon_x$$

$$\alpha_y \sim \mathcal{N}(0, \sqrt{\lambda} \cdot \Sigma)$$

$$\alpha_x \sim \mathcal{N}(0, \sqrt{\rho} \cdot \Sigma)$$

$$\epsilon_y \sim \mathcal{N}(0, \sqrt{1 - \lambda})$$

$$\epsilon_x \sim \mathcal{N}(0, \sqrt{1 - \rho})$$

where Σ is a correlation matrix proportional to either geographic or linguistic proximities between nations, and λ and ρ are autocorrelation parameters that represent the expected spatial or cultural phylogenetic signal for outcome and predictor variables, respectively. Importantly, in this simulation, we know that there is no direct causal relationship between y and x because we simulate the variables independently. Instead, any relationship between the two variables is merely the result of autocorrelation.

We set the autocorrelation parameters to either 0.2 (weak), 0.5 (moderate), or 0.8 (strong). We simulated 100 datasets for each parameter combination, resulting in 900 datasets. Each dataset had 236 rows representing different nations, with the following associated data for each nation: latitude, longitude, continent (Africa, Asia, Europe, North America, Oceania, or South America), and language family of the nation's majority spoken language (Afro-Asiatic, Atlantic-Congo, Austroasiatic, Austronesian, Eskimo-Aleut, Indo-European, Japonic, Kartvelian, Koreanic, Mande, Mongolic-Khitan, Nilotic, Nuclear Trans New Guinea, Sino-Tibetan, Tai-Kadai, Tupian, Turkic, or Uralic).

With the resulting simulated datasets, we standardised outcome and predictor variables and fitted eleven different models: (1) naive regression without controls, (2) regression with latitude control, (3) regression with longitude control, (4) regression with continent fixed effects, (5) regression with language family fixed effects, (6) regression

employing Conley standard errors with 100 km cutoff, (7) regression employing Conley

standard errors with 1,000 km cutoff, (8) regression employing Conley standard errors with

10,000 km cutoff, (9) Bayesian regression including a Gaussian process over latitudes and

longitudes, (10) Bayesian regression including random intercepts covarying according to

linguistic proximity, and (11) Bayesian regression including both a Gaussian process over

latitudes and longitudes and random intercepts covarying according to linguistic proximity.

Models employing Conley standard errors required latitude and longitude values, and

the cutoffs implied the distance beyond which autocorrelation is assumed to be zero. These

models were fitted using the *conleyreg* R package[73]. Bayesian models were fitted using the

brms R package[69]. Our choice of priors was based on prior predictive simulation. All models

converged normally ($\hat{R} < 1.1$). Across all model types and parameter combinations, we

calculated the false positive rate as the proportion of models that estimated slopes with a

95% confidence / credible interval excluding zero. We calculated 95% bootstrap confidence

intervals for these false positive rates with 1,000 bootstrap iterations.

Reanalyses

We searched the individual analyses from our systematic review for cross-national

correlations with available data. We included only analyses for which we were able to

replicate the original result (i.e. find a cross-national correlation with the same sign and

roughly the same effect size). We restricted our search to one analysis per article, and aimed

for an even number of analyses for both economic development and cultural values studies.

We also ensured that at least one analysis was a multilevel model, with multiple observations

per nation.

The twelve analyses that we settled on[13,14,16,51–59] were mostly bivariate cross-national

correlations, except for two. One analysis[14] additionally controlled for log gross national

income, and another analysis[56] is a multilevel model including random intercepts for nations

474 and several individual-level and national-level covariates (see Model 5 in original paper).

475 Before running any additional models, we pre-registered these twelve analyses on the Open

476 Science Framework on 25th January 2022 (https://osf.io/uywx8/).

477 For each individual analysis, we ran four models: (1) a naive regression replicating the

478 original finding, (2) a regression including a Gaussian process allowing nation random

479 intercepts to covary according to a geographic proximity matrix from latitude and longitude

480 values, (3) a regression including nation random intercepts that covaried according to a

481 linguistic proximity matrix, and (4) a regression including both a geographic Gaussian

482 process and nation random intercepts with linguistic covariance. See Supplementary

483 Methods for full models.

484 We fitted these models using the *brms* R package[69]. Our choice of priors was based on

485 prior predictive simulation. All models converged normally ($\hat{R} < 1.1$), though for some

486 models we resorted to using approximate Gaussian processes[74] to reach convergence.

Reproducibility

488 All data and code are accesible at our Open Science Framework repository

489 (https://osf.io/uywx8/). We used the *targets* R package[75] to create a reproducible data

490 analysis pipeline and the *papaja* R package[76] to reproducibly generate the manuscript.

References

1. Caselli, F. Accounting for cross-country income differences. in (eds. Aghion, P. & Durlauf, S. N.) vol. 1 679–741 (Elsevier, 2005).

2. Austin, K. F. & McKinney, L. A. Disease, war, hunger, and deprivation: A cross-national investigation of the determinants of life expectancy in less-developed and sub-Saharan African nations. *Sociological Perspectives* **55**, 421–447 (2012).

3. Rai, D., Zitko, P., Jones, K., Lynch, J. & Araya, R. Country- and individual-level socioeconomic determinants of depression: Multilevel cross-national comparison. *British Journal of Psychiatry* **202**, 195–203 (2013).

4. Diener, E., Diener, M. & Diener, C. Factors predicting the subjective well-being of nations. in *Culture and well-being: The collected works of Ed Diener* (ed. Diener, E.) 43–70 (Springer Netherlands, 2009). doi:10.1007/978-90-481-2352-0_3.

5. Kirkcaldy, B., Furnham, A. & Siefen, G. The relationship between health efficacy, educational attainment, and well-being among 30 nations. *European Psychologist* **9**, 107–119 (2004).

6. White, C. J. M., Muthukrishna, M. & Norenzayan, A. Cultural similarity among coreligionists within and between countries. *Proceedings of the National Academy of Sciences* **118**, e2109650118 (2021).

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

8. Pollet, T. V., Tybur, J. M., Frankenhuis, W. E. & Rickard, I. J. What can cross-cultural correlations teach us about human nature? *Human Nature* **25**, 410–429 (2014).

9. Benhabib, J. & Spiegel, M. M. The role of human capital in economic development:

Evidence from aggregate cross-country data. *Journal of Monetary Economics* **34**, 143–173 (1994).

10. Comin, D., Easterly, W. & Gong, E. Was the wealth of nations determined in 1000 BC? *American Economic Journal: Macroeconomics* **2**, 65–97 (2010).

11. La Porta, R., Lopez-De-Silanes, F., Shleifer, A. & Vishny, R. W. Legal determinants of external finance. *The Journal of Finance* **52**, 1131–1150 (1997).

12. Sachs, J. D. & Warner, A. M. The curse of natural resources. *European Economic Review* **45**, 827–838 (2001).

13. Fincher, C. L., Thornhill, R., Murray, D. R. & Schaller, M. Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings of the Royal Society B: Biological Sciences* **275**, 1279–1285 (2008).

14. Gelfand, M. J. *et al.* Differences between tight and loose cultures: A 33-nation study. *Science* **332**, 1100–1104 (2011).

15. Hofstede, G. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations.* (Sage Publications, 2001).

16. Inglehart, R. & Baker, W. E. Modernization, cultural change, and the persistence of traditional values. *American Sociological Review* **65**, 19–51 (2000).

17. Awad, E. *et al.* The moral machine experiment. *Nature* **563**, 59–64 (2018).

18. Rhoads, S. A., Gunter, D., Ryan, R. M. & Marsh, A. A. Global variation in subjective well-being predicts seven forms of altruism. *Psychological Science* **32**, 1247–1261 (2021).

19. Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P. & Henrich, J. The Church, intensive

536   kinship, and global psychological variation. *Science* **366**, eaau5141 (2019).

537 20. Thomson, R. *et al.* Relational mobility predicts social behaviors in 39 countries and is

538   tied to historical farming and threat. *Proceedings of the National Academy of*

539   *Sciences* **115**, 7521–7526 (2018).

540 21. Kissling, W. D. & Carl, G. Spatial autocorrelation and the selection of simultaneous

541   autoregressive models. *Global Ecology and Biogeography* **17**, 59–71 (2008).

542 22. Hewlett, B. S., DeSilvestri, A. & Guglielmino, C. R. Semes and genes in africa. *Current*

543   *Anthropology* **43**, 313–321 (2002).

544 23. Nunn, N. & Puga, D. Ruggedness: The blessing of bad geography in Africa. *The Review*

545   *of Economics and Statistics* **94**, 20–36 (2012).

546 24. Tobler, W. R. A computer movie simulating urban growth in the Detroit region.

547   *Economic Geography* **46**, 234–240 (1970).

548 25. Tylor, E. B. On a method of investigating the development of institutions; applied to

549   laws of marriage and descent. *The Journal of the Anthropological Institute of Great*

550   *Britain and Ireland* **18**, 245–272 (1889).

551 26. Naroll, R. Two solutions to Galton's Problem. *Philosophy of Science* **28**, 15–39 (1961).

552 27. Naroll, R. Galton's Problem: The logic of cross-cultural analysis. *Social Research* **32**,

553   428–451 (1965).

554 28. Jarque, C. M. & Bera, A. K. A test for normality of observations and regression

555   residuals. *International Statistical Review / Revue Internationale de Statistique* **55**,

556   163–172 (1987).

557 29. Legendre, P. Spatial autocorrelation: Trouble or new paradigm? *Ecology* **74**, 1659–1673

558 (1993).

559 30. Roberts, J., Seán AND Winters. Linguistic diversity and traffic accidents: Lessons from

560 statistical studies of cultural traits. *PLOS ONE* **8**, 1–13 (2013).

561 31. Lichstein, J. W., Simons, T. R., Shriner, S. A. & Franzreb, K. E. Spatial autocorrelation

562 and autoregressive models in ecology. *Ecological Monographs* **72**, 445–463 (2002).

563 32. Loftin, C. & Ward, S. K. A spatial autocorrelation model of the effects of population

564 density on fertility. *American Sociological Review* **48**, 121–128 (1983).

565 33. Murdock, G. P. & White, D. R. Standard cross-cultural sample. *Ethnology* **8**, 329–369

566 (1969).

567 34. Dow, M. M. & Eff, E. A. Global, regional, and local network autocorrelation in the

568 standard cross-cultural sample. *Cross-Cultural Research* **42**, 148–171 (2008).

569 35. Eff, E. A. Does Mr. Galton still have a problem? Autocorrelation in the standard

570 cross-cultural sample. *World Cultures* **15**, 153–170 (2004).

571 36. Symonds, M. R. E. & Blomberg, S. P. A primer on phylogenetic generalised least squares.

572 in *Modern phylogenetic comparative methods and their application in evolutionary*

573 *biology: Concepts and practice* (ed. Garamszegi, L. Z.) 105–130 (Springer Berlin

574 Heidelberg, 2014). doi:10.1007/978-3-662-43550-2_5.

575 37. Watts, J., Sheehan, O., Bulbulia, J., Gray, R. D. & Atkinson, Q. D. Christianity spread

576 faster in small, politically structured societies. *Nature Human Behaviour* **2**, 559–564

577 (2018).

578 38. Atkinson, Q. D., Coomber, T., Passmore, S., Greenhill, S. J. & Kushnick, G. Cultural

579 and environmental predictors of pre-european deforestation on pacific islands. *PLOS*

580 *ONE* **11**, 1–15 (2016).

39. Kelly, M. Understanding persistence. *CEPR Discussion Paper No. DP15246* (2020).

40. Bromham, L., Hua, X., Cardillo, M., Schneemann, H. & Greenhill, S. J. Parasites and politics: Why cross-cultural studies must control for relatedness, proximity and covariation. *Royal Society Open Science* **5**, 181100 (2018).

41. Bromham, L., Skeels, A., Schneemann, H., Dinnage, R. & Hua, X. There is little evidence that spicy food in hot countries is an adaptation to reducing infection risk. *Nature Human Behaviour* **5**, 878–891 (2021).

42. Currie, T. E. & Mace, R. Analyses do not support the parasite-stress theory of human sociality. *Behavioral and Brain Sciences* **35**, 83–85 (2012).

43. Passmore, S. & Watts, J. WEIRD people and the Western Church: Who made whom? *Religion, Brain & Behavior* 1–58 (2022) doi:10.1080/2153599X.2021.1991459.

44. *Human development report.* http://hdr.undp.org/en/composite/HDI.

45. Currie, T. E. & Mace, R. Political complexity predicts the spread of ethnolinguistic groups. *Proceedings of the National Academy of Sciences* **106**, 7339–7344 (2009).

46. Conley, T. GMM estimation with cross sectional dependence. *Journal of Econometrics* **92**, 1–45 (1999).

47. Conley, T. G. Spatial econometrics. in *Microeconometrics* (eds. Durlauf, S. N. & Blume, L. E.) 303–313 (Palgrave Macmillan, 2010).

48. McElreath, R. *Statistical rethinking: A Bayesian course with examples in R and Stan.* (CRC Press, 2020).

49. Neal, R. M. Regression and classification using Gaussian process priors. in *Bayesian statistics* (eds. Bernardo, J. M., Berger, J. O., Dawid, A. P. & Smith, A. F. M.) vol. 6

475–501 (Oxford University Press).

50. Villemereuil, P. de & Nakagawa, S. General quantitative genetic methods for comparative biology. in *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice* (ed. Garamszegi, L. Z.) 287–303 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-662-43550-2_11.

51. Beck, T., Demirgäç-Kunt, A. & Levine, R. Law, endowments, and finance. *Journal of Financial Economics* **70**, 137–181 (2003).

52. Beck, T., Demirgäç-Kunt, A. & Levine, R. SMEs, growth, and poverty: Cross-country evidence. *Journal of economic growth* **10**, 199–229 (2005).

53. Bockstette, V., Chanda, A. & Putterman, L. States and markets: The advantage of an early start. *Journal of Economic growth* **7**, 347–369 (2002).

54. Easterly, W. & Levine, R. Tropics, germs, and crops: How endowments influence economic development. *Journal of Monetary Economics* **50**, 3–39 (2003).

55. Easterly, W. Inequality does cause underdevelopment: Insights from a new instrument. *Journal of Development Economics* **84**, 755–776 (2007).

56. Skidmore, M. & Toya, H. Do natural disasters promote long-run growth? *Economic Inquiry* **40**, 664–687 (2002).

57. Adamczyk, A. & Pitt, C. Shaping attitudes about homosexuality: The role of religion and cultural context. *Social Science Research* **38**, 338–351 (2009).

58. Alesina, A., Giuliano, P. & Nunn, N. On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics* **128**, 469–530 (2013).

59. Knack, S. & Keefer, P. Does social capital have an economic payoff? A cross-country

625 investigation. *The Quarterly Journal of Economics* **112**, 1251–1288 (1997).

626 60. Sookias, R. B., Passmore, S. & Atkinson, Q. D. Deep cultural ancestry and human

627 development indicators across nation states. *Royal Society Open Science* **5**, 171411

628 (2018).

629 61. Smith, M. D., Rabbitt, M. P. & Coleman- Jensen, A. Who are the world's food insecure?

630 New evidence from the Food and Agriculture Organization's Food Insecurity

631 Experience Scale. *World Development* **93**, 402–412 (2017).

632 62. Hastie, T. J. & Tibshirani, R. J. *Generalized additive models.* (Routledge, 2017).

633 63. Smaldino, P. E. & McElreath, R. The natural selection of bad science. *Royal Society*

634 *Open Science* **3**, 160384 (2016).

635 64. Liddell, T. M. & Kruschke, J. K. Analyzing ordinal data with metric models: What could

636 possibly go wrong? *Journal of Experimental Social Psychology* **79**, 328–348 (2018).

637 65. Nicholson, J. S., Deboeck, P. R. & Howard, W. Attrition in developmental psychology: A

638 review of modern missing data reporting and practices. *International Journal of*

639 *Behavioral Development* **41**, 143–153 (2017).

640 66. Tipton, E., Pustejovsky, J. E. & Ahmadi, H. Current practices in meta-regression in

641 psychology, education, and medicine. *Research Synthesis Methods* **10**, 180–194 (2019).

642 67. Chambers, C. D. & Tzavella, L. The past, present and future of Registered Reports.

643 *Nature Human Behaviour* **6**, 29–42 (2021).

644 68. Stan Development Team. RStan: The R interface to Stan. (2020).

645 69. Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *Journal*

646 *of Statistical Software* **80**, 1–28 (2017).

647 70. Oishi, S. Socioecological psychology. *Annual Review of Psychology* **65**, 581–609 (2014).

648 71. Tiefelsdorf, M. & Griffith, D. A. Semiparametric filtering of spatial autocorrelation: The
649   eigenvector approach. *Environment and Planning A: Economy and Space* **39**,
650   1193–1221 (2007).

651 72. De Hoyos, R. E. & Sarafidis, V. Testing for cross-sectional dependence in panel-data
652   models. *The Stata Journal* **6**, 482–496 (2006).

653 73. Düben, C. *conleyreg: Estimations using Conley standard errors.* (2021).

654 74. Solin, A. & Särkkä, S. Hilbert space methods for reduced-rank Gaussian process
655   regression. *Statistics and Computing* **30**, 419–446 (2020).

656 75. Landau, W. M. The targets R package: A dynamic Make-like function-oriented pipeline
657   toolkit for reproducibility and high-performance computing. *Journal of Open Source*
658   *Software* **6**, 2959 (2021).

659 76. Aust, F. & Barth, M. *papaja: Create APA manuscripts with R Markdown.* (2020).

Supplementary Material

Supplementary Methods

Calculating global geographic and linguistic proximity matrices. Geographic distance between two nations was calculated as the logged geodesic distance between country capital cities (data from the R package *maps*; Brownrigg, 2018) using the R package *geosphere* (Hijmans, 2019). The geographic proximity matrix was computed as one minus the log geographic distance matrix scaled between 0 and 1.

Linguistic proximity between two nations was calculated as the cultural proximity between all languages spoken within those nations, weighted by speaker percentages. We acquired cultural proximity data by combining the language family trees provided by Glottolog v3.0 (Hammarström et al., 2017) into one global language tree (undated and unresolved). We calculated cultural proximity s between two languages j and k as the distance (in number of nodes traversed) of their most recent common ancestor i to the root of the tree, through the formula:

$$s_{jk} = \frac{n_r - n_i}{n_r}$$

where n_r is the maximum path length (in number of nodes traversed) leading to the pan-human root r, and n_i is the maximum path length leading to node i. We then combined these proximities with speaker data from Ethnologue 21 (Ethnologue, 2018) and compared every language spoken within those nations by at least 1 permille of the population, weighted by speaker percentages, through the formula:

$$w_{lm} = \Sigma\Sigma p_{lj} p_{mk} s_{jk}$$

678   where p_{lj} is the percentage of the population in nation l speaking language j, p_{mk} is

679 the percentage of the population in nation m speaking language k, and s_{jk} is the proximity

680 measure between languages j and k (Eff, 2008). The resulting linguistic proximity matrix

681 was also scaled between 0 and 1 before analysis.

682   **Bayesian models for reanalysis.**   We provide model formulae for our reanalyses of

683 cross-national correlations, for a general bivariate case with standardised outcome Y and

684 predictor X variables. In the naive regression model without controls for non-independence:

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta X_i$$

$$\alpha \sim \text{Normal}(0, 0.4)$$

$$\beta \sim \text{Normal}(0, 0.4)$$

$$\sigma \sim \text{Exponential}(5)$$

685 The priors in this model were arrived at by prior predictive checks, with wider priors making

686 predictions beyond the scale of standardised outcome variables and narrower priors being too

687 informative.

₆₈₈   To control for spatial non-independence, we add a Gaussian process to this model and
₆₈₉ feed it a scaled geographic distance matrix D based on Euclidean distances between latitude
₆₉₀ and longitude coordinates. This distance matrix is computed internally by the R package
₆₉₁ *brms* (Bürkner, 2017). The Gaussian process uses an exponentiated quadratic covariance
₆₉₂ kernel, the only covariance kernel currently supported by *brms*. The model formula is:

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \kappa_{\text{NATION}[i]} + \beta X_i$$

$$\begin{pmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ ... \\ \kappa_n \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ ... \\ 0 \end{pmatrix}, \mathbf{K} \right)$$

$$\mathbf{K}_{ij} = sdgp^2 \exp\left(-D_{ij}^2/(2lscale^2) \right)$$

$$\alpha \sim \text{Normal}(0, 0.4)$$

$$\beta \sim \text{Normal}(0, 0.4)$$

$$\sigma \sim \text{Exponential}(5)$$

$$sdgp \sim \text{Exponential}(5)$$

$$lscale \sim \text{InverseGamma}(?, ?)$$

₆₉₃ where n is the number of nations, and D_{ij}^2 reflects the squared Euclidean distances between
₆₉₄ latitude and longitude coordinates for the i-th and j-th nations. Notice that the inverse
₆₉₅ gamma prior on *lscale* is left undetermined. This is because the *brms* package intelligently
₆₉₆ tunes the prior for this parameter based on the covariates of the Gaussian process (see
₆₉₇ https://betanalpha.github.io/assets/case_studies/gp_part3/part3.html).

698   To control for cultural phylogenetic non-independence, we manually specify the

699 covariance structure for nation random intercepts using a pre-computed linguistic proximity

700 matrix L (see previous section). The covariance between two nations is assumed to be

701 linearly proportional to the linguistic proximity between those nations. This assumption is

702 justified if we assume that cultural traits evolve via Brownian motion along a language

703 phylogeny. The non-centered parameterisation of this model is:

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + z_{\text{NATION}[i]}\sigma_\alpha L + \beta X_i$$

$$\alpha \sim \text{Normal}(0, 0.4)$$

$$\beta \sim \text{Normal}(0, 0.4)$$

$$z_j \sim \text{Normal}(0, 1)$$

$$\sigma_\alpha \sim \text{Exponential}(5)$$

$$\sigma \sim \text{Exponential}(5)$$

704   Finally, we can control for spatial and cultural phylogenetic non-independence

705 simultaneously by including both a Gaussian process over latitude and longitude coordinates

706 *and* nation random intercepts that covary according to linguistic proximity. The resulting

707 model is as follows:

$$Y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \kappa_{\text{NATION}[i]} + z_{\text{NATION}[i]}\sigma_\alpha L + \beta X_i$$

$$\begin{pmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ ... \\ \kappa_n \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ ... \\ 0 \end{pmatrix}, \mathbf{K} \right)$$

$$\mathbf{K}_{ij} = sdgp^2 \exp\left(- D_{ij}^2/(2lscale^2) \right)$$

$$\alpha \sim \text{Normal}(0, 0.4)$$

$$\beta \sim \text{Normal}(0, 0.4)$$

$$z_j \sim \text{Normal}(0, 1)$$

$$\sigma_\alpha \sim \text{Exponential}(5)$$

$$\sigma \sim \text{Exponential}(5)$$

$$sdgp \sim \text{Exponential}(5)$$

$$lscale \sim \text{InverseGamma}(?, ?)$$

708 **Supplementary Results**

709 **Geographic and cultural phylogenetic signal estimates for Human**

710 **Development Index and Inglehart's value dimensions.** Controlling for shared

711 cultural ancestry, the proportion of national-level variance explained by spatial proximity

712 was 0.37 for the Human Development Index (95% credible interval [0.19 0.60]; $BF_{\neq 0} > 100$),

713 0.43 for traditional values (95% CI [0.19 0.73]; $BF_{\neq 0} > 100$), and 0.18 for survival values

714 (95% CI [0.01 0.44]; $BF_{\neq 0} = 1.77$). Controlling for spatial proximity, the proportion of

715 variance explained by shared cultural ancestry was 0.61 for the Human Development Index

716 (95% CI [0.39 0.80]; $BF_{\neq 0} > 100$), 0.55 for traditional values (95% CI [0.26 0.79]; $BF_{\neq 0} >$

717 100), and 0.79 for survival values (95% CI [0.54 0.96]; $BF_{\neq 0} > 100$).
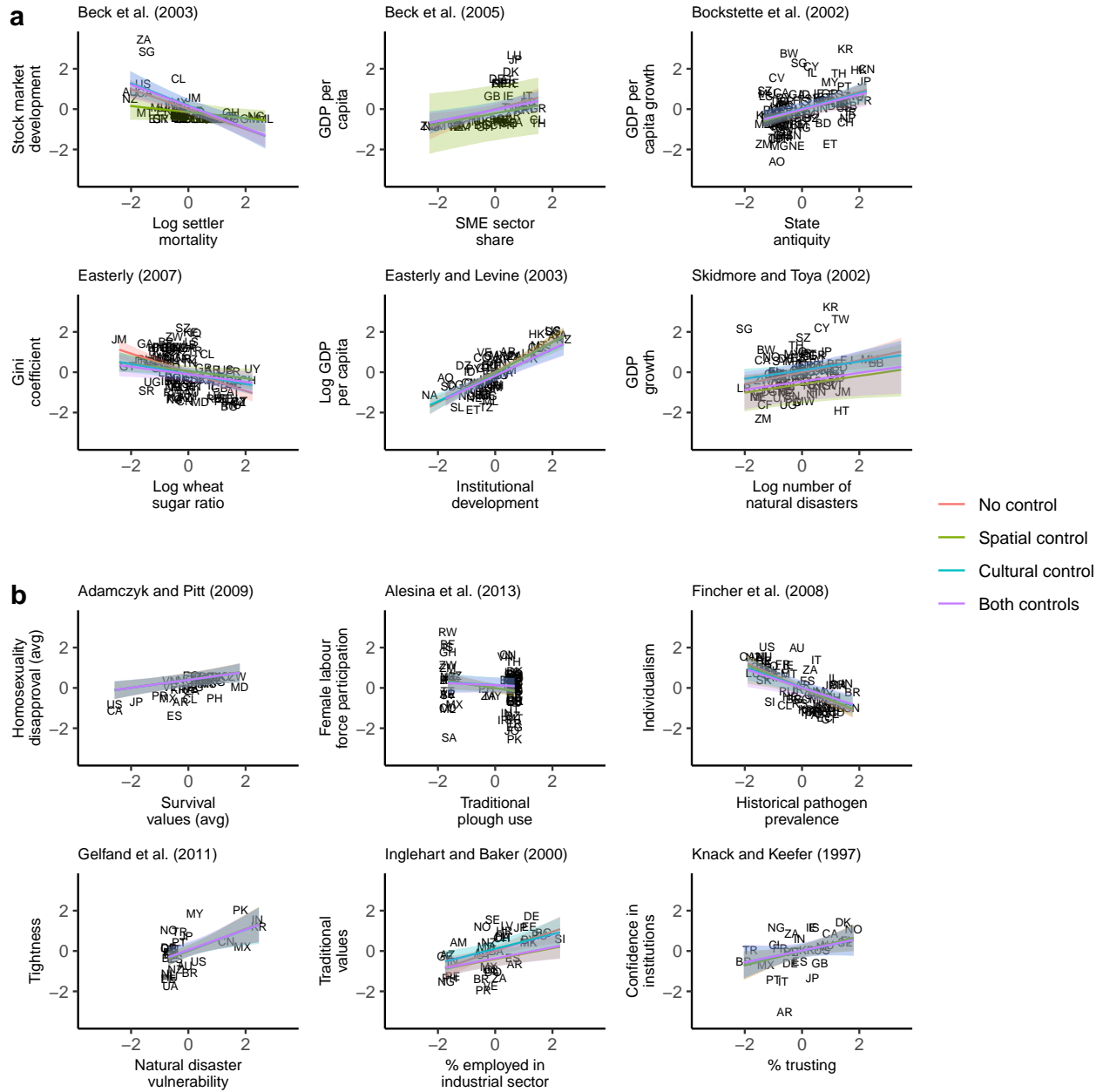
718 **Supplementary Figures**
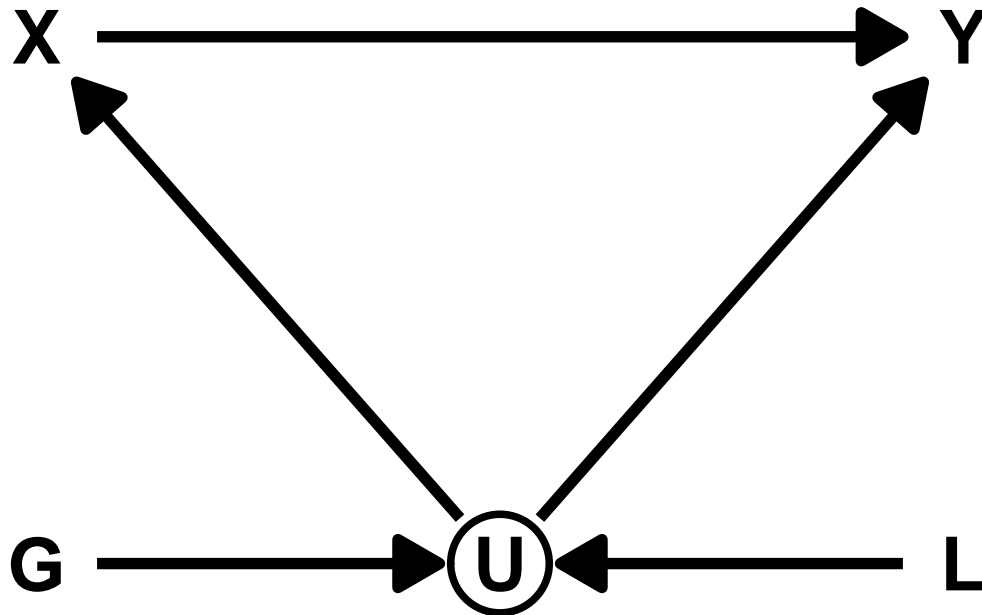
Supplementary Figure S1. A causal directed acyclic graph of spatial and cultural phylogenetic non-independence in cross-national studies. We are interested in estimating the direct effect of national-level exposure X on national-level outcome Y. But these variables are confounded by their common unobserved cause U. U is a stand-in for shared environmental, ecological, and geographic causes (e.g. climate, biodiversity, physical topography) and cultural and institutional causes (e.g. cultural norms, political systems). In this causal model, we need to condition on U to estimate the direct path from X to Y, but we cannot since it is unobserved. However, geographic G and linguistic L relationships between societies influence U, since changing a nation's spatial distance to or shared cultural ancestry with other nations will change its environmental and cultural traits. We can thus use G and L to model the covariation between X and Y induced by U. Failing to do this and simply estimating the bivariate correlation between X and Y will produce spurious relationships and residuals that are spatially and culturally non-independent around the world.
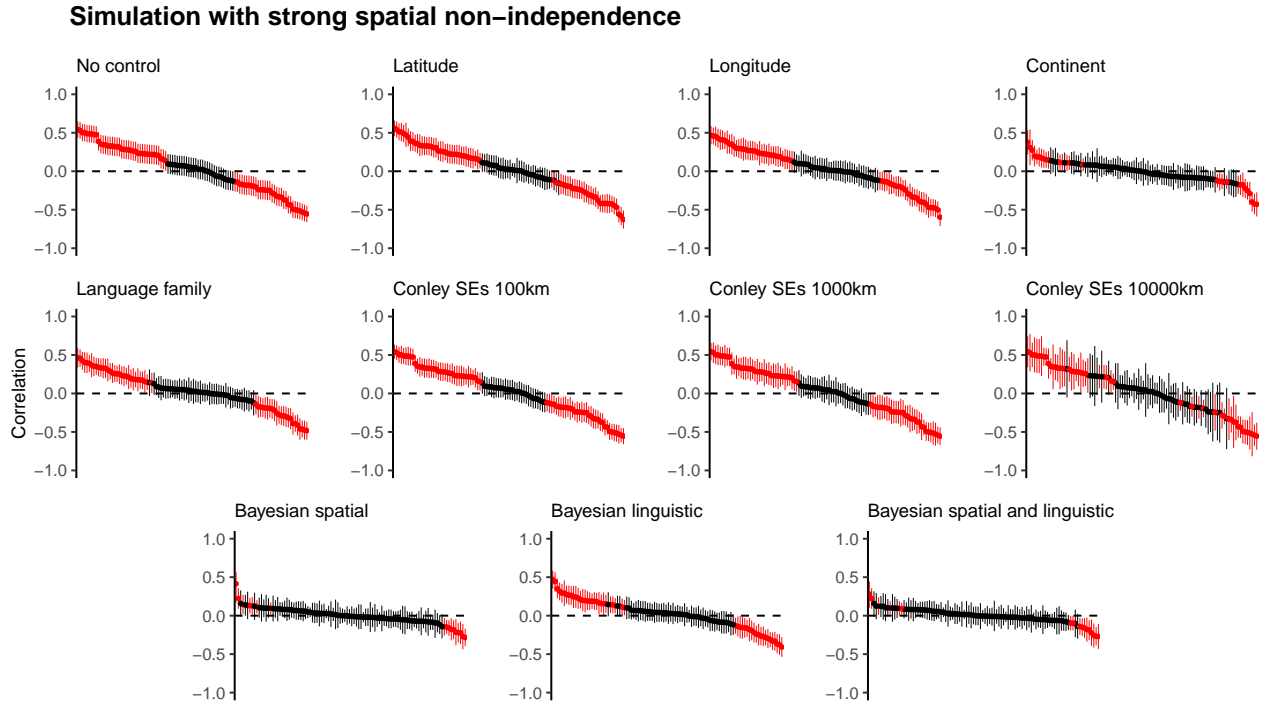
Supplementary Figure S2. Distribution of cross-national correlations from simulation study under strong spatial autocorrelation. In these simulations, the strength of spatial autocorrelation is set to 0.8 for both outcome and predictor variables. For frequentist regression models, point ranges represent correlation estimates and 95% confidence intervals. For Bayesian regression models, point ranges represent posterior means and 95% credible intervals. Correlations are ordered by effect size independently in each panel. Red point ranges indicate that the slope is "significant" (i.e. the 95% confidence / credible interval excludes zero). Black point ranges indicate that the slope is "not significant".

**Simulation with strong cultural phylogenetic non-independence**
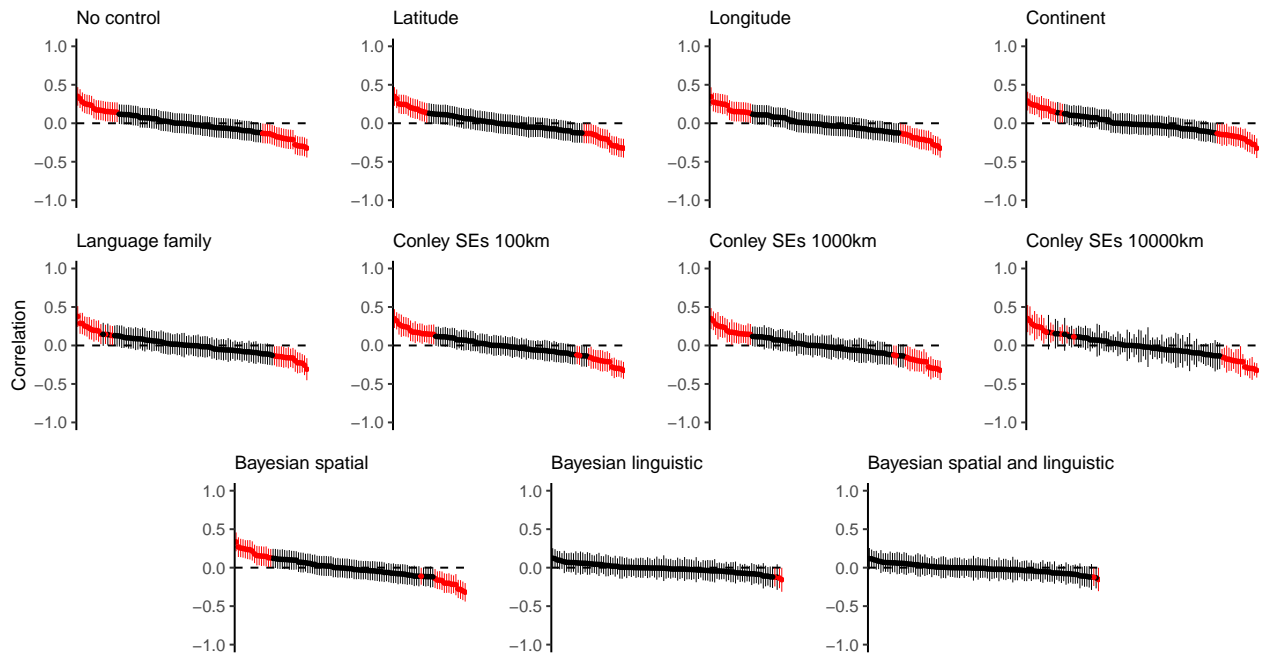
Supplementary Figure S3. *Distribution of cross-national correlations from simulation study under strong cultural phylogenetic autocorrelation.* In these simulations, the strength of cultural phylogenetic autocorrelation is set to 0.8 for both outcome and predictor variables. For frequentist regression models, point ranges represent correlation estimates and 95% confidence intervals. For Bayesian regression models, point ranges represent posterior means and 95% credible intervals. Correlations are ordered by effect size independently in each panel. Red point ranges indicate that the slope is "significant" (i.e. the 95% confidence / credible interval excludes zero). Black point ranges indicate that the slope is "not significant".

Supplementary Figure S4. Reanalysis models fitted to raw data, for economic development (a) and cultural values (b) studies. Data points are labelled using ISO 3166-1 alpha-2 letter country codes. In all reanalyses, outcome and predictor variables are standardised, making regression slopes comparable to Pearson's correlation coefficients. Lines and shaded areas represent posterior median regression lines and 95% credible intervals. For models with covariates (Adamzyck and Pitt 2009; Gelfand et al. 2011), marginal effects are presented holding all covariates at zero or their reference categories.
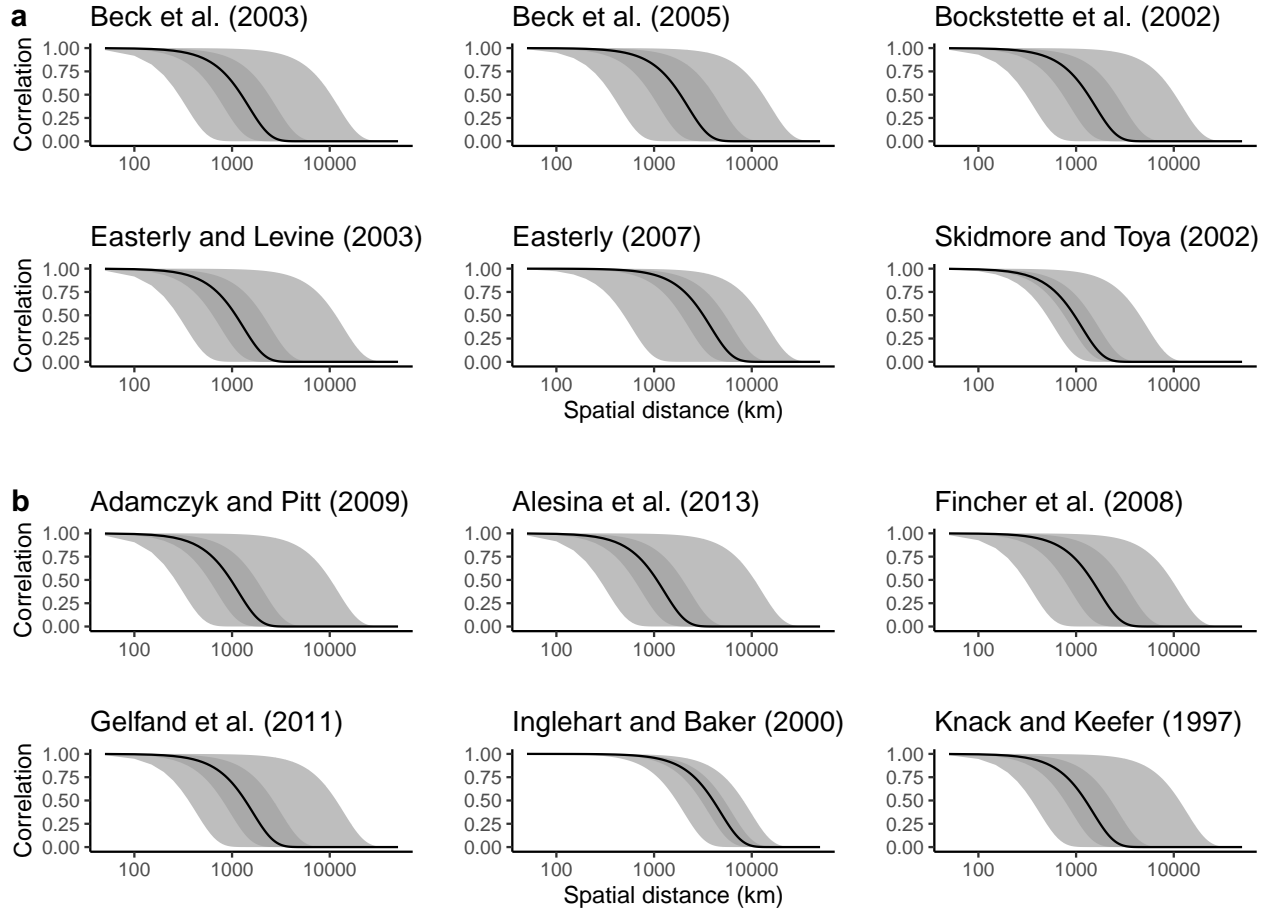
Supplementary Figure S5. Posterior estimates of Gaussian process functions mapping spatial autocorrelation onto geographic distance from our reanalyses of economic development (a) and cultural values (b) studies. Estimates are from models additionally controlling for cultural phylogenetic non-independence. The y-axis represents the amount of spatial autocorrelation between data points with increasing distance between those points on the x-axis (logged distance in kilometres). Lines and shaded areas represent median posterior spatial autocorrelation functions and 50% and 95% credible intervals.
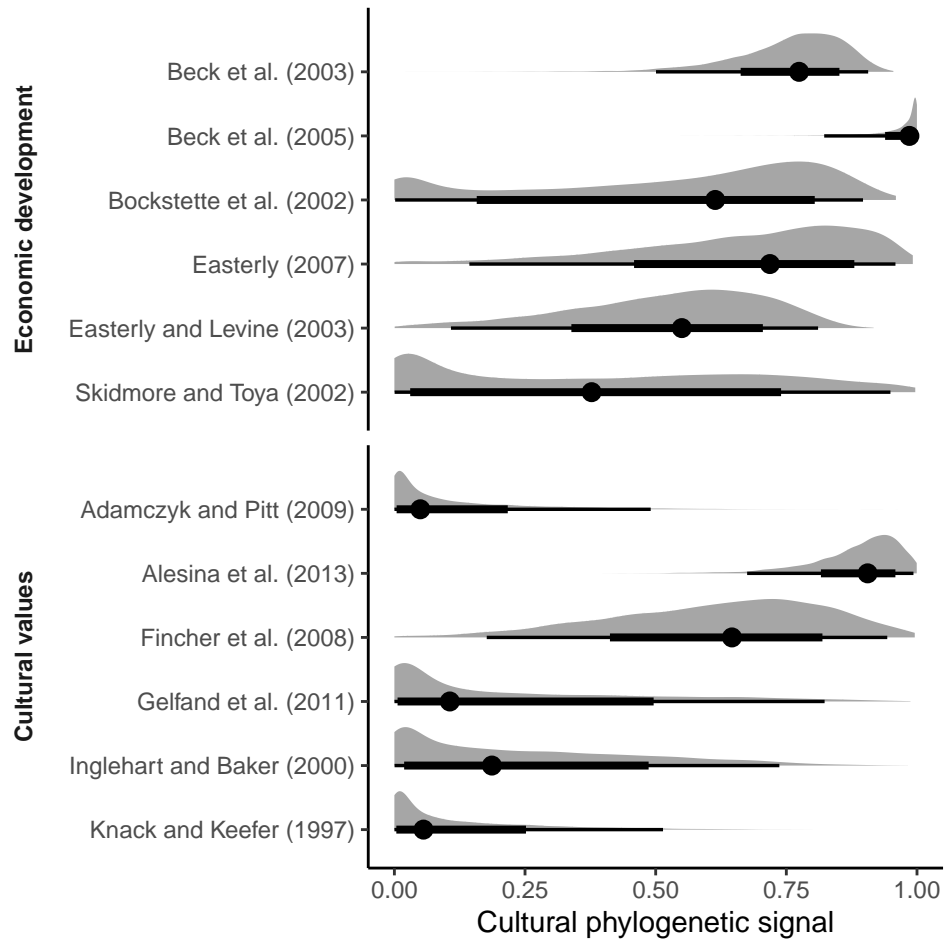
Supplementary Figure S6. Posterior estimates of cultural phylogenetic signal from our reanalyses. Estimates are from models additionally controlling for spatial non-independence. Cultural phylogenetic signal is operationalised as the proportion of national-level variance explained by linguistic proximity between nations. Ridges are full posterior distributions, and points are posterior medians, and lines represent 50% and 95% credible intervals.

Supplementary References

Brownrigg, R. *maps: Draw geographical maps.* (2018).

Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* **80**, 1—28 (2017).

Eff, E. A. Weight matrices for cultural proximity: Deriving weights from a language phylogeny. *Structure and Dynamics* **3**, (2008).

Ethnologue: Languages of the world. (SIL International, 2018).

Hammarström, H., Forkel, R., Haspelmath, M. & Bank, S. *Glottolog 3.0.* (Max Planck Institute for the Science of Human History, 2017). doi:%5B10.5281/zenodo.4061162%5D(https://doi.org/10.5281/zenodo.4061162)

Hijmans, R. J. *geosphere: Spherical trigonometry.* (2019).