






ARTICLE



<https://doi.org/10.1057/s41599-020-00691-9>

OPEN

Predictive modeling of religiosity, prosociality, and moralizing in 295,000 individuals from European and non-European populations

Pierre O. Jacquet^{1,2,4} , Farid Pazhoohi^{3,4}, Charles Findling¹, Hugo Mell^{1,3}, Coralie Chevallier¹  & Nicolas Baumard² 

Why do moral religions exist? An influential psychological explanation is that religious beliefs in supernatural punishment is cultural group adaptation enhancing prosocial attitudes and thereby large-scale cooperation. An alternative explanation is that religiosity is an individual strategy that results from high level of mistrust and the need for individuals to control others' behaviors through moralizing. Existing evidence is mixed but most works are limited by sample size and generalizability issues. The present study overcomes these limitations by applying *k*-fold cross-validation on multivariate modeling of data from >295,000 individuals in 108 countries of the World Values Surveys and the European Value Study. First, this methodology reveals no evidence that European and non-European religious people invest more in collective actions and are more trustful of unrelated conspecifics. Instead, the individuals' level of religiosity is found to be weakly but positively associated with social mistrust and negatively associated with the production of behaviors, which benefit unrelated members of the large-scale community. Second, our models show that individual variation in religiosity is well explained by the interaction of increased levels of social mistrust and increased needs to moralize other people's sexual behaviors. Finally, stratified *k*-fold cross-validation demonstrates that the structures of these association patterns are robust to sampling variability and reliable enough to generalize to out-of-sample data.

¹LNC2, Département d'études cognitives, École normale supérieure, INSERM, PSL Research University, 75005 Paris, France. ²Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University, 75005 Paris, France. ³Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada. ⁴These authors contributed equally: Pierre O. Jacquet, Farid Pazhoohi. ✉email: pierre.ol.jacquet@gmail.com; nbaumard@gmail.com

Introduction

The existence of moral religions is a complex phenomenon whose complete understanding requires explanations at different interacting levels: historical, ecological, economic, social, and psychological (Baumard and Boyer, 2013; Baumard et al., 2015; Berger, 1981; Botero et al., 2014; McKay and Whitehouse, 2015; Norris and Inglehart, 2011; Weber, 1958; Whitehouse et al., 2019). Among the psychological explanations, an influential one is that moralizing religions sustain large-scale cooperation (Norenzayan et al., 2016), i.e., a set of behaviors directed toward unrelated individuals and which provide direct or indirect benefits to both the recipients and the producers (West et al., 2007). The belief in invisible, rule-enforcing agents would increase the believers' compliance with cooperative norms. These cooperative behaviors would in turn increase tolerance, social trust and cooperation between unrelated members of the group, and allow human populations to live in increasingly complex social structures (Henrich et al., 2010; Norenzayan et al., 2016; Turchin, 2011; Wilson, 2003). Ultimately, religious groups would be favored over non-religious groups, leading to the diffusion of moralizing religions across the globe (Johnson and Bering, 2006; Norenzayan et al., 2016; Purzycki et al., 2016).

An alternative explanation is that higher social mistrust should lead individuals to higher religiosity through increased moralization of other people's behaviors (Baumard and Chevallier, 2015; Pazhoohi et al., 2017; Weeden et al., 2008; Weeden and Kurzban, 2013). In this perspective, religiosity is a way for individuals to police others so that they adopt behaviors that are more favorable to their own fitness and their own strategy. Weeden and colleagues (Weeden et al., 2008; Weeden and Kurzban, 2013) for instance, have put forward the "Reproductive Religiosity Model" in which the primary function of religiosity is to make others adopt behaviors that are more favorable to monogamous strategies. Monogamy and high parental investment require increased commitment from marriage partners. Religious beliefs and norms might help increase the effectiveness of moralizing sexual promiscuity, hence making monogamy a more stable strategy. Such policing behaviors should be higher when individuals perceive their environment as non-cooperative, that is, when they have low level of trust in others.

Despite a range of empirical works, these two psychological explanations remain hard to disentangle (see for instance, the recent controversy around Whitehouse and colleagues (Whitehouse et al., 2019)). One of the reasons is the lack of quantitative data, which drastically limits the use of sophisticated statistical tools. Even in modern societies, previous papers have often used very limited samples.

In this paper, we use the European Value Study (<http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp>) and the World Value Survey (<https://dbk.gesis.org/dbksearch/sdesc2.asp?no=4804&db=e&doi=10.4232/1.12253>) (Inglehart et al., 2014), two very large datasets containing information about the values, the beliefs, and the behaviors of >295,000 individuals from high income but also middle-income and low-income countries, including African, Asian, American, European and Oceanian countries. The EVS and the WVS datasets allow to study people's values and beliefs at the individual level. So far, large-scale studies on the origins of religiosity have limited their analyses to the aggregated level (Norris and Inglehart, 2011) (e.g., comparing the average religiosity level between different countries). One major problem with this approach is that it runs the risk of falling into the well-known ecological fallacy. The ecological fallacy occurs when inferences about the nature of individuals are deduced from inferences about the group to which those individuals belong. For instance, the well-known association between religiosity and low levels of GDP per capita at the aggregate level does not tell us whether the same association is valid at the individual level.

A frequently used alternative consists in mixing both aggregated and individual levels (Ruck et al., 2018; Weeden and Kurzban, 2013). However, a second problem comes out once these two levels of description are put together: in the present case, an infinite number of aggregated factors might compete to explain the individual data. Which aggregated factor must be selected to explain the associations between social distrust, cooperation, policing needs, and religiosity? Should it be the countries' geographical vicinity? Their linguistic vicinity? Their population density? Their GDP per capita? Their education rate? Their political system? Any other forms of cultural norms? In any case, the final choice is necessarily arbitrary.

A final issue characterizing existing studies is the systematic lack of generalizability attempts. The capacity of statistical models to generalize their predictions to unknown, out-of-sample data is almost never tested. This weakness considerably limits the scope of their findings.

To overcome these problems, we apply stratified k -fold cross-validation to multivariate structural equation models. By repeatedly testing the models' predictive accuracy on individual data picked-up at random within the European and non-European countries composing the datasets, cross-validation offers a powerful way to extract the behavioral patterns that generalize across individuals regardless of arbitrary country-level factors.

Our methodology consists in few steps. The full dataset is first randomly partitioned into tenfolds of nearly equal size. Subsequently, ten iterations of training and validation are performed such that within each iteration a different fold of the data is held-out for validation (the test data: here representing 10% of the whole sample) while the remaining $k-1$ folds are used for training (the training data, here representing 90% of the whole sample). At each iteration, the covariance matrix fitted by the multivariate structural equation model on the training data is compared to the covariance matrix observed in the test data. The predictive accuracy of the training model is further checked by comparing the fitted parameters on a covariance matrix observed in a test dataset whose individual values have been randomly permuted. The aim is to verify that the model parameters fitted on the training data fail at predicting the test data when its internal structure is broken down by means of random permutations. Therefore, this procedure allows us to fully confirm that the latent structure hypothesized by our structural equation model is present or absent in the data. In the end of the ten iterations, the whole sample is reshuffled and re-stratified before a new round of ten iterations starts. The whole procedure is repeated 100 times (100 rounds of 10 iterations) in order to obtain reliable performance estimation.

Two series of structural equation modeling (SEM) models fitted on individual data were run on each dataset (EVS and WVS). The first class of models test Hypothesis 1 according to which religiosity leads individuals to higher levels of social trust and large-scale cooperation (models 1). The second class of models test Hypothesis 2 according to which higher social mistrust leads individuals to higher religiosity through increased moralization of other people's behaviors. This second class involves models focusing on attitudes aiming to moralize sexual promiscuity to specifically expand on the predictions derived from the "Reproductive religiosity hypothesis" (for a review, see Moon et al., 2019) (models 2.1), and models focusing on attitudes aiming to moralize free-riding (models 2.2).

Methods

EVS and WVS respondents. The EVS and the WVS are independent large-scale sociological surveys exploring people's values and beliefs across countries and times. The EVS and the WVS have relied on social scientists since 1981 to collect data from large representative samples of the populations from >100 countries in total.

Our two hypotheses are first tested on individuals who participated in the 3rd and the 4th waves of the European Value Survey (EVS—data collected between 1999 and 2004), hence taken as the datasets of reference. Unlike the EVS waves 1 and 2, the 3rd and 4th EVS waves allow a reasonably complete modeling of our theoretically specified latent constructs *Religiosity*, *Social mistrust*, *Large-scale cooperation*, *Moralizing sexual promiscuity*, and *Moralizing free-riding*. We also try to replicate the effects eventually found with the EVS models on an independent dataset including individuals from non-European countries, i.e., the World Value Survey (WVS). Multivariate analyses were ran on the WVS waves 3 to 6. For a matter of data incompleteness, the WVS waves 1 and 2 were not included in the analyses.

The EVS dataset includes 164,997 respondents, 107,406 of them being part of the 3rd and 4th EVS waves. Among these 107,406, 100,599 respondents belonging to 46 countries remained after removing those with too many missing values (+ 2SDs from the sample mean), and those who presented missing values in continuous variables that could not be faithfully imputed (i.e., age). The list of countries (and their respective sample size) included in the final EVS dataset is available in the Supplementary Material.

The WVS dataset includes 348,532 respondents, 236,378 of them remained after removing European countries, and 219,533 after removing respondents with too many missing values (+2SDs from the sample mean) and those who presented missing values in continuous variables that could not be faithfully imputed (i.e., age). Finally, for a matter of data completeness, we focused our analyses on the WVS waves 3 to 6, ending up with a final sample of 195,914 respondents belonging to 62 countries. The list of countries (and their respective sample size) included in the final WVS dataset is available in the Supplementary Material.

Missing data. Multiple imputation techniques were used to preserve sample size and avoid biased estimations of model parameters. Twenty complete EVS datasets and 20 complete WVS datasets were generated by fully conditional specifications for categorical, ordinal and continuous data (i.e., predictive mean matching, logistic regression imputation, proportional odds model). The reader can find in the SM Tables S1 and S2 the percentages of imputed values by items, for each dataset.

Multivariate analyses. The EVS and WVS datasets were independently subjected to multivariate analyses through SEM. Structural equation models involve two major parts: a “measurement” model and a “structural” model. The measurement model consists in relating a number of directly observed variables—labeled “indicators”—to a smaller set of hypothesized constructs—labeled “latent” variables—presumed to cause the covariations among the indicators, hence reflecting a continuum that is not directly observable. The measurement model allows theoretical specification of latent variables as a priori hypotheses to be tested against the observed covariations of the data. The a priori assignment of each indicator to the theoretically specified latent variables leads to the reduction of the number of parameters estimated by the model, which is then expected to enhance the efficiency of parameter estimation. The “structural” model consists in estimating a restricted set of a priori specified pathways between the latent variables identified by the measurement model. Therefore, SEM allows for the modeling of the covariations among all these factors by a theoretically driven model (Kline, 2015), a procedure particularly adapted for the purpose of the present study. For each dataset, three structural equation models were designed: model 1 tested the Hypotheses 1, while model 2.1 and model 2.2 tested the Hypothesis 2.

Specification of the measurement model. For the EVS and the WVS models, the measurement part aims to represent *Religiosity*, *Social mistrust*, *Large-scale cooperation*, *Moralizing sexual promiscuity*, and *Moralizing free-riding* as latent variables from several indicators. While the EVS dataset contains enough items to allow the modeling of *Social mistrust* as a latent construct as well, in the WVS data the number of items, which could reflect interpersonal mistrust is smaller and the item reflecting general mistrust only provides a binary response. To overcome this limitation, we chose to model *Social mistrust* from the WVS data as a single composite variable (more details below). The full list of EVS and WVS indicators as well as their original scores are reported in SM Tables S1 and S2, and in the Supplementary Methods section. In addition, descriptive statistics of the indicators involved in the EVS and WVS models 1, 2.1, and 2.2 are available in the SM Tables S3.

Specification of the structural model. The structural part of the EVS and the WVS models 1 allows testing our hypothesis 1 by estimating two regression parameters representing the following direct associations: (i) *Religiosity* and *Large-scale cooperation*, (ii) *Religiosity* and *Social mistrust*. A covariance term was added in order to estimate the correlations between *Large-scale cooperation* and *Social mistrust*. Let’s now focus on the EVS and the WVS models 2.1 and 2.2. Their structural part allows testing our hypothesis 2 by estimating three regression parameters representing the following direct associations: (i) *Social mistrust* and *Religiosity*, (ii) *Social mistrust* and *Moralizing sexual promiscuity/free-riding*, (iii) *Moralizing sexual promiscuity/free-riding* and *Religiosity*. In addition, the models allows the estimation of the indirect effect of *Social mistrust* on *Religiosity* via *Moralizing sexual promiscuity/free-riding*.

Religiosity

EVS models 1, 2.1, and 2.2. The latent variable reflecting the individuals’ level of *Religiosity* was modeled from six indicators (SM Table S1) covering various aspects like ritualistic practices, commitment to religious faith, confidence into religious institutions, or beliefs in religious moral.

WVS models 1, 2.1, and 2.2. Since religious moral beliefs item are not available in the WVS data, *Religiosity* is modeled as a latent variable involving the first 5 *Religiosity* items (WVS code: F028, E069_01, F034, A006, F063) used in the EVS models (SM Table S2).

Social mistrust

EVS models 1, 2.1, and 2.2. *Social mistrust* was modeled as a latent variable aiming to capture the covariations of two indicators (SM Table S1). The 1st indicator is labeled General mistrust, and expresses the sum of scores (z-transformed and then rescaled between 0 and 1) on the three general trust items classically reported in the social sciences literature. The 2nd indicator takes the form of a single composite variable aiming to quantify the number of social categories (up to 15) that the respondents would not like to have as neighbors. We labeled it the Interpersonal mistrust indicator, by opposition to the General mistrust indicator. Scores on the Interpersonal mistrust indicator is expressed in terms of the sum of positive answers on the 15 available social categories.

WVS models 1, 2.1, and 2.2. In the WVS models, *Social mistrust* was measured as a single composite variable instead of a latent construct. The reason is that in the WVS, the only available general trust item involves a binary response. Since it is often difficult to handle binary measures with structural equation models, which also involve ordinal and continuous measures (Kline, 2015), we opted for summing the binary General mistrust

(z -) scores and the Interpersonal mistrust (z -) scores described above (SM Table S2). Note that in the WVS database, six social categories (instead of 15 in the EVS database) could be exploited to compute the Interpersonal trust score (i.e., “People of different race”, “Heavy drinkers”, “Immigrants”, “People who have AIDS”, “Drug addicts”, “Homosexuals”). The sum of the general trust and the Interpersonal trust z -scores was then rescaled between 0 and 1. Greater scores on this single composite variable, therefore, indicate greater levels of mistrust.

Large-scale cooperation

EVS model 1. Large-scale cooperation was modeled as a latent variable aiming to capture the covariations of 6 indicators (SM Table S1). The 1st *Large-scale cooperation* indicator is Volunteering in collective activities, a single composite variable aiming to quantify the number of social groups or organizations for which the respondents freely dedicate time and energy without receiving any financial compensation (up to 15). The other five indicators reflect the respondents’ involvement in several types of political actions. Item scores were coded such that greater scores indicate greater involvement in each type of political actions. Importantly, the Volunteering in collective activities indicators and the five Involvement in political actions indicators offer the advantage of tapping actual behaviors involving tangible costs for the agent and concrete benefits for others (Lettinga et al., 2020). Although these two measures are limited in their own way (for instance, many cooperative interactions happen outside the activities described in these items), they are the most concrete instances of true investment in collective goods available in the EVS and the WVS.

WVS model 1. In the WVS sample, only three political actions indicators are available. The items used in the EVS to reconstruct the respondents’ volunteering in collective activities are unfortunately empty of data. Therefore, *Large-scale cooperation* was modeled as a latent construct involving the 3 political actions indicators (SM Table S2).

Moralizing sexual promiscuity

EVS model 2.1. A number of recent studies suggest that religiosity might be primarily associated with a willingness to moralize behaviors deviating from sexual monogamy, but less with free-riding (for a recent review, see Moon et al., 2019). In line with these works, we model moralizing attitudes separately for promiscuous sexual behaviors and free-riding. *Moralizing sexual promiscuity* was hence modeled as a latent variable aiming to capture the covariations of 6 indicators (SM Table S1) reflecting the tolerance of the respondents when confronted to several practices. The six indicators’ scores were coded such that greater scores indicate greater intolerance towards promiscuous sex.

WVS model 2.1. In the WVS dataset, the approach used to model moralizing attitudes towards sexual promiscuity was almost identical to the one described above (SM Table S2). The only difference is that in the WVS model, the *Moralizing sexual promiscuity* latent variable could only be captured from 4 instead of 6 indicators.

Moralizing free-riding

EVS model 2.2. Moralizing free-riding was modeled as a latent variable aiming to capture the covariations of seven indicators (SM Table S1) reflecting the tolerance of the respondents when confronted to several free-riding behaviors. The seven indicators’ scores were coded such that greater scores indicate greater intolerance towards free-riding.

WVS model 2.2. In the WVS dataset, the *Moralizing free-riding* latent variable only involves three instead of seven indicators (SM Table S2).

Covariates. The respondents’ age at the time of the interview was used as the adjustment variable in each of the EVS and WVS models. Two reasons explain this choice. First, including too many covariates in a model often limits the interpretability of the results (Pearl, 2009). Second, it decreases the parsimony of the model by artificially inflating its predictive value with factors, which are somehow minimally relevant with regard to the purpose of the study. Introducing the age of the respondents as the adjustment variable is nevertheless necessary as its effects on religiosity (Ruck et al., 2018), prosocial behaviors (Freund and Blanchard-Fields, 2014) or moralizing attitudes (Truett, 1993) are well-documented. Each effect estimated by the models was, therefore, adjusted for the respondents’ age.

Assessing models’ fit. To overcome issues of sample size dependency of the χ^2 fit metrics, together with the non-normality of our datasets (see SM Tables S3 for descriptive statistics), we opted for a weighted-least squares estimator (WLSMV) and report the χ^2 , *CFI*, *RMSEA*, and *SRMR* statistics corrected by a scaling factor, which compensates for the average kurtosis of the data (Satorra and Bentler, 2001). A model’s fit is generally considered excellent when the *RMSEA* is close to 0.05, the *CFI* close to 0.95 and *SRMR* close to 0.08.

Fit indices and parameter estimates. The fit indices and parameter estimates reported in the main text and in the SM are the average of the 1000 iterations that have been set for cross-validating a model. The scaled versions of the chi-squared, the *CFI*, the *RMSEA* (and the 95% confidence intervals) and of the *SRMR* statistics are labeled as follows: χ^2 .scaled, *CFI*.scaled, *RMSEA*.scaled [CI.lower; CI.upper], *SRMR*.scaled. The unstandardized values of the indicators’ loadings composing the latent variables, the regression coefficients characterizing the associations between the latent/composite variables, and the residual covariance of variables are labeled *P.unstd*. The standardized versions of these parameters are labeled *P.std*. We also report under the *z* label the value of the Wald test (and its *p*-value)—the ratio between the parameter value and its standard error—as an indicator of the simple effect size and statistical significance. The observed and model-implied (fitted) correlation matrices averaged on the 1000 iterations of each model can be found at the end of the Supplementary Materials, SM Tables S6, S7 and S8.

Effect sizes. In the SEM literature, effect sizes can be estimated in different ways. The 1st way is to refer to the *z*-values and the associated standard deviation of each parameter estimation (the ratio between the *z*-values and *sd* gives an idea of an effect’s power), as well as the parameter’s confidence intervals. Effect sizes can also be appreciated from a parameter’s standardized coefficient (Niemenen et al., 2013). A standardized coefficient refers to how many standard deviations the dependent variable will change per a standard deviation increase in the independent variable, holding all other variables constant. The standardized coefficients of a regression path linking two latent variables can eventually be interpreted similar to a correlation, where <0.2 is considered a weak effect, between 0.2 and 0.5 as a moderate effect and >0.5 as a strong effect (Acocck, 2014). The same logic can be applied to factor loadings, which might be represented as the correlation between the latent variable and the observed indicators. Finally, the square of factor loadings and path coefficients (similar to the coefficient of determination R^2) can be used to calculate the proportion of variance of

the observed indicator that is explained by the latent variable or, alternatively, the proportion of variance of the latent variable that is captured by the predictors.

However, the size of an effect is relative. Its scope depends on the theory that motivates the search for an association between the variables of interest. For example, one may run a regression model to study the association between the temperatures measured at each hour of a day by two different marketed thermometers placed at the very same location. Let imagine that the obtained standardized regression coefficient is 0.6. According to the benchmarks reported above this is a strong effect. However, one has a theory about how the two variables are expected to be linked: measurements made by two marketed thermometers must be extremely correlated. According to this theory, a standardized coefficient of 0.6 is an extremely disappointing effect size, and it certainly indicates that there is a device that measures anything but temperature.

Testing the predictive accuracy of the models. It is of course unrealistic to hypothesize such a strong association between behavioral/psychological constructs in general and, in the more specific case of the present study, between religiosity, large-scale cooperation, moralizing needs or social trust. Even if these associations are weak or moderate, the most important thing is to assess their robustness and reliability. This is precisely the reason why we used stratified cross-validation. Cross-validation is the ultimate analytic layer that allows to test whether the observed effects are robust to sampling variability on the one hand, and reliable enough to generalize to out-of-sample data on the other hand.

In addition, cross-validation is especially relevant in the present context because the inherent complexity of SEMs (high number of fitted parameters) often increases the risk of overfitting. Overcoming this risk and ensuring the capacity of a SEM to predict new, out-of-sample data is, therefore, necessary to ensure its validity (MacCallum et al., 1992). Yet, this task is more than rarely achieved in practice (Breckler, 1990; Whittaker and Stapleton, 2006). The present study takes up this challenge by using a stratified k -fold method (Arlot and Celisse, 2010). It allows to assess the predictive performance of the models and to judge how they perform outside the sample to a new dataset. The motivation to use cross-validation techniques is that when a model is fitted, it is fitted to a training dataset. Without cross-validation one can only have information on how does the model perform to the in-sample data. Ideally, one would like to see how the model perform in terms of accuracy of its predictions when one has a new dataset. In brief, for a given model:

1. The full dataset (either the EVS or the WVS datasets, depending on the model) is first randomly partitioned into tenfolds of nearly equal size.
2. Subsequently ten iterations of training and validation are performed such that within each iteration a different fold of the data is held-out for validation (the *test data*: here representing 10% of the whole sample) while the remaining $k-1$ folds are used for fitting (the *training data*, here representing 90% of the whole sample).
3. At each iteration, the model-implied covariance matrix estimated from the training data is compared to the covariance matrix directly observed in the test data. The smaller the difference between the two matrices, the greater the predictive power of the training model. Matrix comparison is achieved by means of a weighted-least square discrepancy function F , defined as a weighted sum of squared residuals (Browne and Cudeck, 1992). To calculate F , we use the weight matrix exploited by the WLSMV estimator to fit the training dataset. A F -value of 0 indicates that the model and parameter estimates perfectly reproduce

the real data. Hence, the closer to zero the F -value, the greater the predictive power of the model. For facilitating the interpretation, the F -value is converted into a χ^2 metric, which is then used to calculate the absolute fit index $RMSEA$ (using the degree of freedom of the test model and the sample size of the test dataset) (Chen, 2007).

4. The predictive value of the training model is further checked by applying the same cross-validation procedure on a test dataset whose individual values have been randomly permuted *within* each indicator. The aim is to check that the fitted values obtained from modeling the training data failed at predicting the test data when its internal structure is broken down by means of randomization. This procedure ultimately tests whether the latent structure hypothesized by our structural equation model is present or absent in the training samples. A *predictive accuracy ratio* is calculated by dividing the F -values obtained from the cross-validation procedure applied on the *randomly permuted* test data by the F -values obtained from the cross-validation procedure applied on the *real* test data. A ratio equal to 1 indicates that the predictive accuracy of the model is null, because it fits the *real* and the *randomly permuted* test datasets equally well. The more the ratio deviates positively from 1, the greater the predictive accuracy of the model. The mean and standard deviation of this ratio, therefore, provide standardized performance estimations allowing a direct comparison of the models' predictive accuracy.
5. At the end of the ten iterations, the whole sample is reshuffled and re-stratified before a new round of ten iterations starts.
6. The whole procedure is repeated 100 times (100 rounds of 10 iterations) in order to obtain reliable performance estimation (the mean and standard deviations across iterations of the predictive accuracy ratio). Note that at each iteration within a round, the training and the test datasets always include different data points.

Analytic plan. The analyses therefore comprise: (i) the EVS and the WVS models each fitted on 1000 training sets; (ii) stratified k -fold cross-validation analyses testing the capacity of the training models to predict the test data; (iii) stratified k -fold cross-validation analyses ensuring that the training models are unable to predict randomly permuted test data. All statistical analyses were carried out in R (<https://www.r-project.org/>) with R Studio. Missing data of the EVS and the WVS datasets were imputed using the R package *mice* (van Buuren and Groothuis-Oudshoorn, 2011). The EVS and WVS models were fitted using the R package *lavaan* (Rosseel, 2012) and the function *runMI* of the R package *semTools* (Jorgensen et al., 2018; Li et al., 1991) was used to pool the parameter estimates, the standard errors and the fit indices obtained for the 20 imputed datasets. The WLSMV estimator was used for its robustness to deviations from normality.

Results

Descriptive statistics of the EVS and WVS variables involved in the models, as well as their observed correlations can be accessed through SM Tables S3, S5, S6, and S7 of the Supplementary Materials. In order to improve the reading of the text, all statistical indices and parameter estimates corroborating the description of the models are reported in the tables included in the main text and in the Supplementary Materials. The models' goodness of fit indices are reported in Table 1. The parameter estimates of their measurement and structural parts are reported in Table 2 for the EVS dataset, and in SM Table S4 for the WVS dataset. The coefficients of determination R^2 are reported in Table 3 as

Table 1 Goodness of fit indices.

	EVS Model 1	EVS Model 2.1	EVS Model 2.2
<i>n</i> respondents	90,539	90,539	90,539
<i>n</i> parameters	46	46	49
df	74	74	87
χ^2 .scaled	8966.958	18,409.525	9102.514
CFI.scaled	0.974	0.956	0.970
RMSEA.scaled	0.036	0.052	0.034
[ci.lower—ci. upper]	[0.036–0.037]	[0.052–0.053]	[0.033–0.034]
SRMR.scaled	0.033	0.037	0.029
	WVS Model 1	WVS Model 2.1	WVS Model 2.2
<i>n</i> respondents	1,76,323	1,76,323	1,76,323
<i>n</i> parameters	30	33	33
df	25	33	33
χ^2 .scaled	8817.046	12,204.92	4097.663
CFI.scaled	0.963	0.956	0.981
RMSEA.scaled	0.045	0.046	0.026
[ci.lower—ci. upper]	[0.044–0.045]	[0.045–0.046]	[0.026–0.027]
SRMR.scaled	0.028	0.030	0.022

The fit indices reported in the table are averaged on the 1000 iterations that have been set for cross-validating each model. A RMSEA \leq 0.05, a CFI \geq 0.95 and a SRMR \leq 0.08 indicate an excellent fit.

indices of effect sizes (i.e., strength of the associations between latent variables and their indicators, strength of the associations between latent variables). Finally, cross-validation results of are reported in Table 4.

Goodness of fit indices. All the EVS and WVS models provide an excellent fit. This is supported by scaled CFI values $>$ 0.95, scaled RMSEA values $<$ 0.05, and scaled SRMR values $<$ 0.08 (Table 1). These results indicate that the discrepancies between the sample and the model-implied covariance matrices are minimal (RMSEA), and that the models performed better than their *null* versions including only the variance of the indicators as parameters (CFI, SRMR).

Measurement models. The measurement parts of these models show that all the indicators load positively and significantly on their respective hypothetic latent variable (EVS: Table 2, Fig. 1a, b, c; WVS: SM Table S4, SM Fig. S1.a, S1.b, S1.c).

First, respondents who score high on the latent *Religiosity* attend religious services more often, place a higher confidence in religious institutions, define themselves as religious persons and, in the case of the EVS respondents specifically, place a higher value in religion in general and in god in particular and display a greater number of religious moral beliefs. The coefficients of determination R^2 reported in Table 3 and estimated in all the EVS models showed that the latent *Religiosity* explained between 46 and 77% of the indicators' variance. In the WVS models, the percentage of the indicators' variance explained by the latent *Religiosity* was comprised between 37 and 67%.

Second, the EVS respondents who score high on the latent *Social mistrust* declare mistrusting other people more and are more inclined to reject members of specific social categories as potential neighbors. However, as shown in Table 3, the proportion of explained variance is rather small for the *General mistrust* indicator (between 11 and 15% as a function of the model), and moderate for the *Interpersonal mistrust* indicator (between 26 and 0.33% as a function of the model).

Third, respondents scoring high on the latent *Large-scale cooperation* undertake more actions that benefit the large-scale

community. More specifically, they are more likely to sign petitions, to join boycotts, to attend lawful demonstrations, to join unofficial strikes and, in a weaker extent, to occupy buildings or factories. This is supported by an explained variance comprised between 22 and 53% in the EVS model 1, and between 40 and 58% in the WVS model 1 (Table 3). However, the EVS respondents who score high on the latent *Large-scale cooperation* are only slightly more inclined than others to volunteer for collective activities. This is reflected by the fact that only 3% of variance of this indicator is captured by the latent variable (see Table 3).

Fourth, respondents who score high on the latent *Moralizing sexual promiscuity* express greater needs of moralizing homosexuality, prostitution, abortion, divorce and, in the case of the EVS respondents specifically, greater needs of moralizing adultery and casual sex. As reported in Table 3, the latent variable accounts for 22 to 61% of the indicators' variance in the EVS model 2.1, and for 44 to 63% of the indicators' variance in the WVS model 2.1.

A similar pattern is finally observed with the latent *Moralizing free-riding*. Respondents scoring high in this latent express greater needs of moralizing individuals who claim government benefits while they are not eligible to, who avoid a fare on public transport, who cheat on taxes, who accepts a bribe and, in the case of the EVS respondents specifically, individuals who steal a car (joyriding), those who lie, and those who pay cash in order to avoid taxes. In the EVS model 2.2, the latent variable *Moralizing free-riding* captures 22 to 61% of the indicators' variance while in the WVS model 2.2 this proportion is comprised between 44 and 56% (Table 3).

Structural models. The Hypothesis 1 states that religiosity should lead individuals to higher levels of trust and large-scale cooperation. The structural parts estimated by the EVS (Table 1a and Fig. 1a) and the WVS models 1 (SM Table S4.a and SM Fig. S1.a) show the opposite. In the EVS model 1, respondents high on the latent *Religiosity* are also high on the latent *Social mistrust*, while they are low on the latent *Large-scale cooperation*. The same pattern emerges from the WVS model 1. Even though these associations are weak (the percentage of variance in *Social mistrust* and *Large-scale cooperation* explained by *Religiosity* is respectively of 8 and 4% in the EVS model 1, and 3% in the WVS model 1, see Table 3), they are nevertheless robust to sampling noise injected by the cross-validation procedure. In sum, the respondents who display the highest levels of religiosity are slightly more likely than others to mistrust unfamiliar people, to reject more social categories as potential neighbors, and to undertake less actions benefiting the large-scale community. Note that in both models, *Social mistrust* and *Large-scale cooperation* negatively covary: respondents displaying the highest levels of mistrust show the lowest level of large-scale cooperation.

We now turn to the structural parts of the EVS model 2.1 (Table 1b and Fig. 1b) and the WVS model 2.1 (SM Table S4.b and SM Fig. S1.b). In line with the Hypothesis 2, respondents high on the latent *Social mistrust* are also high on the latent *Moralizing sexual promiscuity*. On the other hand, those with higher scores on *Moralizing sexual promiscuity* also display higher levels of *Religiosity*. Importantly, the EVS model 2.1 reveals no direct effect of *Social mistrust* on *Religiosity*. Instead, the effect of *Social mistrust* is fully transmitted to *Religiosity* via the *Moralizing sexual promiscuity* latent variable. The mediation effect of the latent *Moralizing sexual promiscuity* is still present in the WVS model 2.1, though in a weaker extent. Overall, the EVS model 2.1 and the WVS model 2.1 remarkably accounted for 28 and 17% of the variance in *Religiosity*, respectively (Table 3).

The structural patterns estimated by the models 2.1 are not entirely reproduced by the EVS model 2.2 (Table 1c and Fig. 1c), nor by the WVS model 2.2 (SM Table S4.b and SM Fig. S1.b). While respondents showing higher scores on the latent

Table 2 Measurement and structural parts estimated by the EVS model 1, 2.1, and 2.2.

(a) Model 1. Parameter estimates and statistics (averaged on the 1000 iterations set for cross-validation)

Model part	Latent variables	Indicators	P.unstd	se	z	p-value	ci.lower	ci.upper	P.std
Measurement model	Religiosity	=~ Religious service attendance	1.681	0.006	277.226	<0.001	1.668	1.692	0.697
		=~ Confidence in religious institutions	0.705	0.003	255.538	<0.001	0.699	0.711	0.715
		=~ Religious person	0.412	0.002	215.694	<0.001	0.409	0.416	0.705
	Social mistrust	=~ Important in life: religion	0.828	0.003	324.027	<0.001	0.823	0.833	0.789
		=~ Important in life: god	2.841	0.007	391.227	<0.001	2.827	2.855	0.868
		=~ Moral beliefs	1.240	0.004	283.374	<0.001	1.231	1.248	0.680
		=~ General mistrust	0.092	0.026	67.710	<0.001	1.707	1.808	0.520
		=~ Interpersonal mistrust	1.757	0.001	63.888	<0.001	0.089	0.095	0.385
		=~ Involvement in collective activities	0.196	0.005	41.208	<0.001	0.187	0.206	0.169
	Large-scale cooperation	=~ Signing petitions	0.535	0.002	223.677	<0.001	0.530	0.540	0.665
		=~ Joining boycotts	0.458	0.002	198.937	<0.001	0.453	0.462	0.714
		=~ Attending lawful demonstrations	0.535	0.002	229.333	<0.001	0.529	0.539	0.722
		=~ Joining unofficial strikes	0.307	0.002	139.224	<0.001	0.302	0.311	0.566
Structural model	Large-scale cooperation	=~ Occupying building or factories	0.191	0.002	95.856	<0.001	0.187	0.195	0.442
		~ Religiosity	0.304	0.007	41.371	<0.001	0.290	0.319	0.291
Covariance	Social mistrust	~ Religiosity	-0.213	0.004	-47.658	<0.001	-0.222	-0.204	-0.208
		-- Large-scale cooperation	-0.389	0.007	-53.700	<0.001	-0.404	-0.375	-0.389

(b) Model 2.1 Parameter estimates and statistics (averaged on the 1000 iterations set for cross-validation)

Model part	Latent variables	Indicators	P.unstd	se	z	p-value	ci.lower	ci.upper	P.std
Measurement model	Social mistrust	=~ General mistrust	0.095	0.001	68.002	<0.001	0.092	0.098	0.381
		=~ Interpersonal mistrust	1.856	0.025	76.466	<0.001	1.809	1.903	0.526
	Moralizing sexual promiscuity	=~ Homosexuality	1.917	0.013	152.402	<0.001	1.893	1.942	0.672
		=~ Prostitution	1.212	0.010	118.904	<0.001	1.193	1.233	0.589
		=~ Abortion	1.976	0.013	157.012	<0.001	1.952	2.001	0.771
		=~ Divorce	1.824	0.012	154.021	<0.001	1.801	1.847	0.723
		=~ Adultery	0.791	0.008	94.875	<0.001	0.775	0.808	0.448
	Religiosity	=~ Having casual sex	1.311	0.011	123.066	<0.001	1.290	1.332	0.584
		=~ Religious service attendance	1.435	0.006	236.518	<0.001	1.423	1.447	0.703
		=~ Confidence in religious institutions	0.600	0.003	223.309	<0.001	0.595	0.606	0.719
		=~ Religious person	0.344	0.002	195.211	<0.001	0.340	0.347	0.694
		=~ Important in life: religion	0.707	0.003	269.159	<0.001	0.702	0.713	0.796
		=~ Important in life: god	2.394	0.008	293.586	<0.001	2.378	2.410	0.864
Structural model	(a) Moralizing sex. prom.	=~ Moral beliefs	1.045	0.004	237.976	<0.001	1.034	1.054	0.677
		~ Social mistrust	0.635	0.011	55.252	<0.001	0.612	0.657	0.534
	(b) Religiosity	~ Moralizing sexual promiscuity	0.528	0.007	77.832	<0.001	0.515	0.541	0.529
		~ Social mistrust	0.007	0.007	0.789	=0.448	-0.010	0.025	0.006
Indirect effect	(c) Religiosity	:= a * b	0.335	0.006	54.488	<0.001	0.323	0.347	0.284

(c) Model 2.2 Parameter estimates and statistics (averaged on the 1000 iterations set for cross-validation)

Model part	Latent	Indicator	P.unstd	se	z	p-value	ci.lower	ci.upper	P.std
Measurement model	Social mistrust	=~ General mistrust	0.087	0.002	44.066	<0.001	0.083	0.090	0.346
		=~ Interpersonal mistrust	2.045	0.043	47.105	<0.001	1.960	2.131	0.579
	Moralizing free-riding	=~ Claiming government benefits	1.010	0.010	102.702	<0.001	0.991	1.030	0.489
		=~ Avoiding a fare on public transport	1.392	0.010	139.174	<0.001	1.373	1.413	0.599
		=~ Cheating on taxes	1.497	0.010	151.943	<0.001	1.478	1.521	0.674
		=~ Someone accepting a bribe	0.963	0.010	102.256	<0.001	0.945	0.982	0.586
		=~ Stealing a car	0.546	0.009	64.622	<0.001	0.529	0.563	0.416
		=~ Lying	1.494	0.009	162.367	<0.001	1.476	1.512	0.684
	Religiosity	=~ Paying cash in order to avoid taxes	1.588	0.010	162.102	<0.001	1.569	1.607	0.625
		=~ Religious service attendance	1.585	0.007	239.409	<0.001	1.572	1.598	0.700
		=~ Confidence in religious institutions	0.659	0.003	221.308	<0.001	0.653	0.665	0.711
		=~ Religious person	0.387	0.002	198.979	<0.001	0.383	0.390	0.703
		=~ Important in life: religion	0.783	0.003	272.909	<0.001	0.777	0.789	0.794

Table 2 (continued)

(c) Model 2.2 Parameter estimates and statistics (averaged on the 1000 iterations set for cross-validation)

Model part	Latent	Indicator	P.unstd	se	z	p-value	ci.lower	ci.upper	P.std
Structural model		=~ Important in life: god	2.666	0.009	304.262	<0.001	2.648	2.683	0.867
		=~ Moral beliefs	1.163	0.005	245.486	<0.001	1.154	1.172	0.679
	(a) Moralizing free-riding	~ Social mistrust	0.030	0.006	4.720	<0.001	0.017	0.042	0.030
	(b) Religiosity	~ Moralizing free-riding	0.290	0.005	45.544	<0.001	0.202	0.220	0.199
	(c) Religiosity	~ Social mistrust	0.290	0.008	38.177	<0.001	0.275	0.305	0.273
	Indirect effect	:= a * b		0.006	0.001	4.823	<0.001	0.004	0.009

(a) EVS model 1, (b) EVS model 2.1, (c) EVS model 2.2. The unstandardized and standardized values of the indicators' loadings composing the latent variables (denoted ==~), the regression coefficients characterizing their associations (denoted ~), and the residual covariances (denoted --) are labeled P.unstd and P.std, respectively. The indirect effects are denoted by the := operator. We also report under the z label values of Wald tests (and their p-values) as indicators of simple effect sizes.

Table 3 Proportion of explained variance of indicators and latent variables (coefficient of determination R²).

		EVS			WVS		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Latent	Religiosity	–	0.28	0.11	–	0.17	0.03
Indicators	Religious service attendance	0.50	0.51	0.51	0.36	0.38	0.37
	Confidence in religious institutions	0.53	0.54	0.52	0.38	0.39	0.38
	Religious person	0.51	0.49	0.51	0.43	0.39	0.44
	Important in life: religion	0.64	0.65	0.65	0.68	0.67	0.66
	Important in life: god	0.77	0.76	0.76	0.53	0.52	0.52
	Moral beliefs	0.46	0.44	0.46	–	–	–
Latent	Social mistrust	0.08	–	–	0.03	–	–
Indicators	General mistrust	0.15	0.14	0.11	–	–	–
	Interpersonal mistrust	0.26	0.27	0.33	–	–	–
Latent	Large-scale cooperation	0.04	–	–	0.03	–	–
Indicators	Involvement in collective activities	0.03	–	–	–	–	–
	Signing petitions	0.44	–	–	0.58	–	–
	Joining boycotts	0.53	–	–	0.50	–	–
	Attending lawful demonstrations	0.53	–	–	0.40	–	–
	Joining unofficial strikes	0.34	–	–	–	–	–
	Occupying building or factories	0.22	–	–	–	–	–
Latent	Moralizing sexual promiscuity	–	0.28	–	–	0.05	–
Indicators	Homosexuality	–	0.47	–	–	0.56	–
	Prostitution	–	0.37	–	–	0.49	–
	Abortion	–	0.61	–	–	0.63	–
	Divorce	–	0.54	–	–	0.44	–
	Adultery	–	0.22	–	–	–	–
	Having casual sex	–	0.40	–	–	–	–
Latent	Moralizing free-riding	–	–	0.00	–	–	0.00
Indicators	Claiming government benefits	–	–	0.25	–	–	0.36
	Avoiding a fare on public transport	–	–	0.40	–	–	0.49
	Cheating on taxes	–	–	0.48	–	–	0.56
	Someone accepting a bribe	–	–	0.36	–	–	0.47
	Stealing a car	–	–	0.18	–	–	–
	Lying	–	–	0.50	–	–	–
	Paying cash in order to avoid taxes	–	–	0.41	–	–	–

The coefficients R² represent the square of factor loadings or path coefficients and are used as measures of effects' sizes. They correspond to the proportion of variance of the observed indicator that is explained by the latent variable or, alternatively, the proportion of variance of the latent variable that is captured by the predictors.

Moralizing free-riding also show elevated levels of *Religiosity*, the consistency of the link between *Social mistrust* and *Moralizing free-riding* is almost null in the EVS model 2.2, and does not reach the standard significance threshold in the WVS model 2.2. Finally, in these models *Social mistrust* has a positive direct effect on *Religiosity* that is not satisfactorily mediated by the *Moralizing free-riding* latent variable. As a result, the percentage of explained variance in *Religiosity* drops to 11% in the EVS model 2.2, and to only 3% in the WVS model 2.2 (Table 3).

Overall, results of the EVS and the WVS models 2.1 remarkably fit the Hypothesis 2. This hypothesis is not fully validated by the EVS and the WVS models 2.2.

Stratified cross-validation. Crucially, all the EVS and the WVS models show good predictive accuracy and, therefore, well generalize to out-of-sample test data ($F \approx 0$; $\chi^2_{\text{model1}} < \chi^2_{\text{null.model1}}$;

$RMSEA < 0.05$) (Table 3). This is further supported by the values of predictive accuracy ratio, which show that the models fail at predicting test data whose internal structure is broken down by means of random permutations (EVS: Fig. 1d; WVS: SM Fig. S1. d). Overall, cross-validation results corroborate the strength and weaknesses of the latent structures established by our models detailed above, as well as the strength and weaknesses of their associations. In addition, a modified Brown-Forsyth Levene-type test of equality of variance comparing the cross-validated distributions of the predictive accuracy ratio of the EVS model 2.1 and 2.2 reveals that the former is less corrupted by sampling noise than the latter (Levene's test statistic = 276.22, $p < 2.2e-16$). This is an indication that the EVS model 2.1 is more robust and reliable than the EVS model 2.2. A similar pattern is observed for the WVS model 2.1, relative to the WVS model 2.2 (Levene's test statistic = 185.62, $p < 2.2e-16$).

Table 4 Stratified tenfolds cross-validation results. Performance estimates (mean ± sd).

	Real test data		Randomly permuted test data		Model comparison	
	F		χ ²		RMSEA	
	Mean (±sd)	Mean (±sd)	Mean (±sd)	Mean (±sd)	Mean (±sd)	Mean (±sd)
EVS	0.103 (+0.007)	1038.677 (+66.548)	7.820 (+0.057)	78 663.257 (+577.933)	0.273 (+0.001)	76.042 (+4.870)
Model 1						
EVS	0.143 (+0.007)	1438.145 (+70.351)	11.054 (+0.069)	111 190.46 (+577.933)	0.324 (+0.001)	77.498 (+3.783)
Model 2.1						
EVS	0.092 (+0.008)	927.762 (+78.875)	7.633 (+0.057)	76 780.612 (+571.348)	0.252 (+0.001)	83.449 (+7.021)
Model 2.2						
WVS	0.041 (+0.002)	796.215 (+49.161)	2.410 (+0.020)	47 204.293 (+384.187)	0.232 (+0.001)	59.511 (+3.693)
Model 1						
WVS	0.043 (+0.002)	837.374 (+45.850)	3.332 (+0.022)	65 284.53 (+429.89)	0.246 (+0.001)	78.196 (+4.291)
Model 2.1						
WVS	0.022 (+0.002)	427.923 (+32.391)	2.120 (+0.017)	41 533.45 (+326.62)	0.196 (+0.001)	97.604 (+7.309)
Model 2.2						

At each iteration of each model, the model-implied covariance matrix estimated from the training data is compared to the covariance matrix directly observed in the test data (real and randomly permuted). Matrix comparison is achieved by means of a weighted-least square discrepancy function *F*. The closer to zero the *F*-value, the greater the predictive power of the model. For facilitating the interpretation of the cross-validation results, the *F*-value is converted into a χ² metrics. This conversion is further used to calculate the absolute fit index RMSEA (using the degree of freedom involved in the training model). Finally, a predictive accuracy ratio is calculated to allow for model comparison. It indicates how much time a model is better at predicting real data than random data.

Discussion

The aim of the present study was to test two psychological hypotheses supposedly explaining the prevalence of moralizing religions in the modern world. The first hypothesis posits that religiosity observed at the individual should be associated with *higher* levels of social trust and *higher* levels of large-scale cooperation (i.e., investments in collective activities and political actions). The second hypothesis posits that *higher* social mistrust should lead individuals to *higher* religiosity through *higher* moralization of other people’s behaviors. Both hypotheses were tested by applying stratified *k*-fold cross-validation on structural equation models fitted on >295,000 participants in the European Value Study and the World Value Survey (Inglehart et al., 2014), and belonging to both Western and non-Western countries. The robustness of the models’ results and their reliability to predict out-of-sample data allow us to disconfirm the Hypothesis 1 and to validate—at least partially—the Hypothesis 2 (i.e., when moralizing needs concern other people’s sexual behaviors).

More specifically, none of the predictions derived from the first hypothesis are verified. Both the EVS model 1 and the WVS model 1 reveal that higher levels of religiosity are robustly associated with lower levels of large-scale cooperation even if this association is weak. Indeed, participants declaring being highly religious invest slightly less in collective activities and participate slightly less to political actions. In addition, they also declare feeling more threatened, whether it is by unrelated people in general or by members of specific social categories. Hence, in both the European and the non-European samples we found no evidence that religiosity is associated with higher level of prosociality, and this might increase policing needs as a result.

This last idea is supported further by the EVS and the WVS models 2.1 showing that a higher level of social mistrust leads to higher religiosity via increased needs to moralize behaviors deviating from sexual monogamy (homosexuality, prostitution, abortion, divorce, adultery, having casual sex). This can be explained by the fact that the most moralizing individuals—at least for the reproductive domain—might not be the most prosocial ones. However, this interpretation does not hold when moralizing needs concern other people’s cooperative behaviors since neither the EVS model 2.2 nor the WVS model 2.2 reveal the hypothesized pattern of associations between social mistrust, moralizing free-riding, and religiosity.

Importantly, we have shown thanks to stratified *k*-fold cross-validation that our models’ results are robust to sampling variability and reliable enough to predict out-of-sample data randomly drawn from both European and non-European samples. In other words, the absence of a negative link between religiosity and social mistrust so as the absence of a positive link between religiosity and large-scale cooperation, are invariably observed from an out-of-sample dataset to another. The same is true for the mediating role of moralizing sexual promiscuity in the positive link between social mistrust and religiosity and, finally, for the null mediating role of moralizing free-riding behaviors in that same link.

More generally, these results converge with a large set of works in the social sciences demonstrating that religiosity is more prevalent in the societies where social trust is lower, legal institutions weaker, welfare state lower and interpersonal violence higher, that is, in societies where large-scale cooperation is *weaker* (Inglehart, 1999; Guiso et al., 2006). Religious beliefs are much higher in Sub-Saharan Africa, in South America or in South Asia than in Western Europe, North America or East Asia (Norris and Inglehart, 2011). Within the US, religious beliefs are also much higher in states where violence, social mistrust and poverty are higher (Norris and Inglehart, 2011). All these observations run against the idea that religion is a factor that

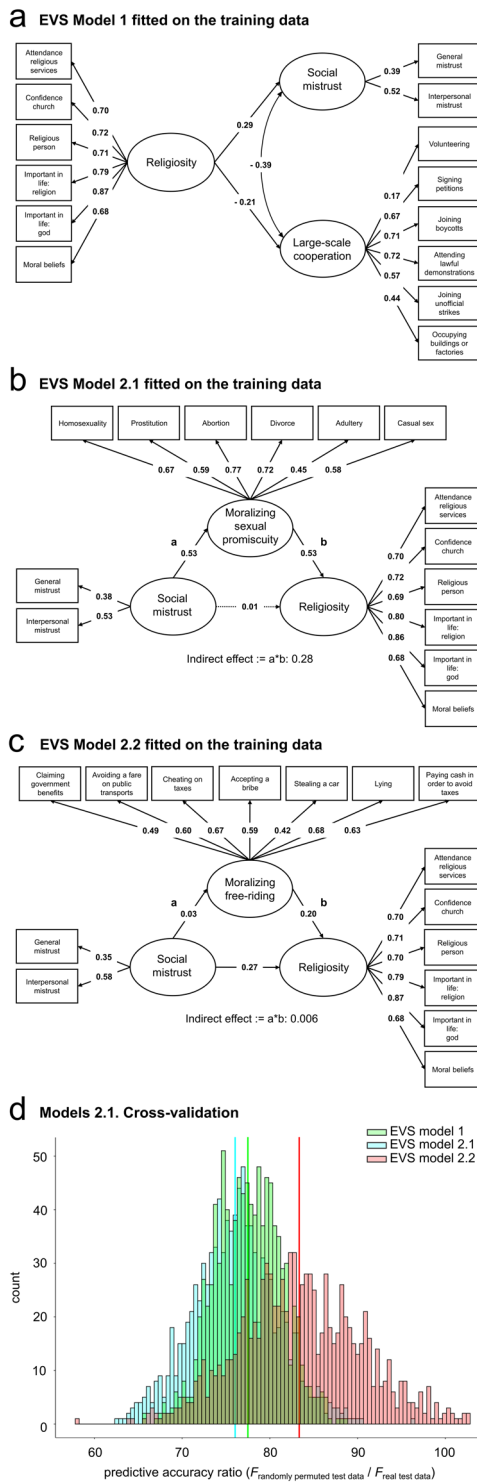


Fig. 1 Measurement and structural models. **a** EVS model 1, **b** EVS model 2.1, **c** EVS model 2.2. Standardized parameter values estimated by the structural equation models applied on the training dataset (90% of the full sample) and averaged on the 1000 iterations set for cross-validation. Ellipses represent latent variables, rectangles represent their indicators. In the WVS model, social mistrust is modeled as a single composite variable, here represented by a rectangle. Paths between the latent variables and/or the single composite variable “Social mistrust” are regressed against the covariate “age of the respondent”. Significant paths at the 5% level are represented with bold arrows. **d** Cross-validation. Distributions of predictive accuracy ratios obtained from the stratified tenfolds cross-validation of the EVS model 1, 2.1 and 2.2 (10*100 rounds in total for each model).

and high in social mistrust that she/he feels the need to influence others. People might therefore engage in religion in the hope of fostering trust and cooperation via moralizing others.

Obviously, despite the methodological strategy we developed to ensure that the observed effects are not false positives, the scope of our findings remains limited on several aspects.

First, their validity depend on the validity of the measures. For example, while indicators used to model *Religiosity* and *Social mistrust* are canonical in the social sciences (Norris and Inglehart, 2011; Inglehart, 1999), those we used to model large-scale cooperation can be subject to discussion. One could argue that they could reflect something else. The reason why we used volunteering in collective activities and involvement in various political actions as indicators of large-scale cooperation is because they represent behaviors having real and tangible costs for the individuals (time and energetic resources are spent without financial pay-offs). From this perspective, they are more relevant than declarative attitudes in the sense that they allow grasping part of an individual’s motivation to pay the cost inherent to large-scale cooperation (Lettinga et al., 2020). Also, we cannot entirely rule out the possibility that the “volunteering” indicator reflects a more general prosocial trait, for example, openness to novelty. More dramatically, engaging in political actions of the kinds reported in the models might reflect anything but prosociality. For instance, a white supremacist might be ready to pay the cost of spending time, money and energy to spread his/her cause and might, therefore, engaged in political actions such like signing petitions, attending demonstrations, etc. We have good hope that it is not the case however. Indeed, the social mistrust indicators (general, interpersonal) negatively correlate with the “political actions” and the “volunteering” indicators (SM Table S5). This leads to a moderate correlation between the *Social mistrust* and the *Large-scale cooperation* latent variables in the EVS model 1 (Table 1a), and a weak correlation in the WVS model 1 (SM Table S4.a). A reverse pattern would have been expected if those behaviors were indeed driven by anti-sociality. We are, therefore, reasonably confident that what is captured by our hypothetical latent constructs is a trait akin to the motivation to cooperate in large social networks.

Second, our models only explain a limited part of variance of religiosity. Our study must therefore be placed in a broader research context, which also incorporates non-psychological, exogenous determinants (e.g., history, ecology, demography, etc.) that could not be taken into account in our models (Baumard and Boyer, 2013; Baumard et al., 2015; Berger, 1981; Botero et al., 2014; McKay and Whitehouse, 2015; Norris and Inglehart, 2011; Weber, 1958; Whitehouse et al., 2019).

Third, the fact that our models do not reveal a positive association between European and non-European people’s level of

contributes to increase social complexity via the promotion of large-scale cooperation.

In sum, what the 1st series of models suggests is that moralizing religiosity and the belief in supernatural punishment is not driven by the need to cooperate with unrelated people nor by the trust individuals place in others. How to make sense of this? An interpretation, albeit counterintuitive at first hand, can be derived from the 2nd series of models: support for moralizing religions and supernatural punishment might be associated with a lower level of large-scale cooperation and a higher level of mistrust because it is precisely when an individual is low in cooperation

religiosity and large-scale cooperation (taking for granted that our measure of cooperation indeed captures effective cooperation, see above) does not mean in any way that people who report being religious never cooperate. In fact, cooperation might arise at different levels of a social structure that the present data does not allow to grasp. It might arise between individuals of a small community, between individuals whose mutual interests are more direct (Santos et al., 2016), between individuals who share parental links (West et al., 2007) or common social values (Handley and Mathew, 2020).

Fourth, one important distinction should be kept in mind, namely the distinction between the religions of small-scale societies, also called “shamanism”, “paganism”, “wild religions”, and the religions of large-scale societies, also called world religions (Boyer and Baumard, 2016; Boyer, 2008; Bellah, 2011; Singh, 2018). The former mostly involve non-moral agents such as ancestors, ghosts, spirits, witches and focus on performing rituals, offering sacrifices, and respecting taboos in order to ward off misfortune and ensure prosperity (Boyer, 2019). By contrast, the latter mostly involve moral agents—such as God, saints, devils, or divine judges—which play a central role, and focus on moral reform and self-discipline (Baumard et al., 2015; Obeyesekere, 2002; Katz, 2008; Segal, 2010). In this perspective, it is important to note that our paper exclusively deal with modern world religions and large-scale cooperation. It cannot address the issue of cooperation and religion in small-scale societies.

Moving back to large-scale societies and moralizing religions, we note that our paper does not explain the ultimate origins of the cultural variability of supernatural beliefs. Why are some countries more religious than others? This paper suggests that religiosity is partly determined by the need to moralize others and ultimately by the level of social trust (i.e., what people think of others’ level of cooperation). We can thus hypothesize that religiosity and the need to moralize others will vary according to the level of cooperation in a group. When cooperation is low, individuals feel more of a need to moralize others; when cooperation is high, individuals feel less of a need to control others.

However, what then explain the cultural variability of cooperation? An emerging body of work suggest that living standard, levels of resources, and environmental harshness partially determined levels of cooperation (Korndörfer et al., 2015; McCullough et al., 2013; Safra et al., 2016; Mell et al., 2019). One likely explanation is that when resources are low and unpredictable, individuals start their life with little capital (embodied capital, economic capital, human capital, etc.) and as such must meet basic needs as adults. Individuals with low capital should be impatient to collect resources such that to improve their productivity and/or survival. In that situation, avoiding waiting costs by favoring short-term rewards is a strategy that might prove advantageous. By contrast, for those who inherited a lot of capital, accumulating new benefits has a marginal value. In that situation being patient and waiting for future and potentially greater pay-offs is a more profitable strategy (Belsky et al., 2012). Since cooperation is a strategy that pay-off on the long run, lower levels of resources are likely to be associated with lower levels of cooperation (Lettinga et al., 2020). Lower levels of cooperation would produce lower levels of social trust, creating a need to moralize others and to promote the belief in supernatural punishment.

This approach is particularly promising for the domain of reproduction. A large body of work has documented the association between harshness and a short-term reproductive decision-making strategy (lower level of sexual commitment, lower level of parental investment) (Belsky et al., 2012; Kusawa et al., 2020). In harsher environment, individuals may thus feel more threatened by the sexual promiscuity of others and more willing to moralize them.

Such a theory would explain the strong association between lower economic development, higher levels of religiosity and

higher level of intolerance often documented in the social sciences literature. Obviously, more work is needed, especially to test the association between religiosity, levels of resources and reproductive strategies at the individual level.

Data availability

The EVS and WVS data used in the present study are respectively available at <http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp> and at <https://dbk.gesis.org/dbksearch/sdesc2.asp?no=4804&db=e&doi=10.4232/1.12253>. The EVS and the WVS data used in the present study, the R codes developed to extract and pre-process the data, to perform multiple imputations, to fit and cross-validate the models on the imputed data files, as well as files containing all the results, are available at <https://osf.io/27rpg/>.

Received: 20 July 2020; Accepted: 26 November 2020;

Published online: 21 January 2021

References

- Acock AC (2014) A gentle introduction to Stata. Stata Press, College Station. Texas
- Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. *Stat Surv* 4:40–79
- Baumard N, Boyer P (2013) Explaining moral religions. *Trends Cogn Sci* 17:272–280
- Baumard N, Chevallier C (2015) The nature and dynamics of world religions: a life-history approach. *Proc Roy Soc B Biol Sci.* 282:20151593
- Baumard N, Hyafil A, Morris I, Boyer P (2015) Increased affluence explains the emergence of ascetic wisdoms and moralizing religions. *Curr Biol* 25:10–15
- Bellah RN (2011) Religion in human evolution. Harvard University Press.
- Belsky J, Schlomer GL, Ellis BJ (2012) Beyond cumulative risk: distinguishing harshness and unpredictability as determinants of parenting and early life history strategy. *Dev Psychol* 48:662–673
- Berger PL (1981) The class struggle in American religion. *Christ Century* 98:194–199
- Botero CA, Gardner B, Kirby KR, Bulbulia J, Gavin MC, Gray RD (2014) The ecology of religious beliefs. *Proc Natl Acad Sci USA* 117:16784–16789
- Boyer P (2008) Religion explained. Random House.
- Boyer P, Baumard N (2016) Diversity of religious systems across history: An evolutionary cognitive approach. In: Liddle JR, Shackelford TK (eds) The oxford handbook of evolutionary psychology and religion. Oxford University Press, Oxford, UK, pp. 1–25
- Boyer P (2019) Informal religious activity outside hegemonic religions: wild traditions and their relevance to evolutionary models. *Religion Brain Behav* 10:459–472
- Breckler SJ (1990) Applications of covariance structure modeling in psychology: Cause for concern? *Psychol Bull* 107:260–273
- Browne MW, Cudeck R (1992) Alternative ways of assessing model fit. *Sociol Methods Res* 21:230–258
- Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Modeling* 14:464–504
- Freund AM, Blanchard-Fields F (2014) Age-related differences in altruism across adulthood: making personal financial gain versus contributing to the public good. *Dev Psychol* 50:1125–1136
- Guiso L, Sapienza P, Zingales L (2006) Does culture affect economic outcomes? *J Econ Perspect* 20:23–48
- Handley C, Mathew S (2020) Human large-scale cooperation as a product of competition between cultural groups. *Nat Comm* 11:702
- Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? *Behav Brain Sci* 33:61–83
- Inglehart R (1999) Trust, well-being and democracy. In: Warren M (ed.) Democracy and trust. Cambridge University Press, Cambridge, pp. 88–120
- In: Inglehart R Haerpfer C Moreno-Alvarez A (eds) (2014) World values survey: all rounds-Country pooled datafile version. JD Systems Institute, Madrid, Spain. Retrieved from www.worldvaluessurvey.org/WVSDocumentationWVL.jsp
- Johnson DDP, Bering JM (2006) Hand of God, mind of man: punishment and cognition in the evolution of cooperation. *Evol Psychol* 4:219–233
- Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y (2018) semTools: Useful tools for structural equation modeling. R package version 0.5-0. Retrieved from <https://CRAN.R-project.org/package=semTools>
- Katz PR (2008) Divine justice: religion and the development of Chinese legal culture. Routledge

- Kline RB (2015) Principles and practice of structural equation modeling, fourth edition. Guilford Publications (2015).
- Korndörfer M, Egloff B, Schmukle SC (2015) A large scale test of the effect of social class on prosocial behavior. *PLoS ONE* 10:e0133193
- Kusawa CW, Adair L, Bechayda SA, Borja JRB, Carba DB, Duazo PL, Eisenberg DTA, Georgiev AV, Gettler LT, Lee NR, Quinn EA, Rosenbaum S, Rutherford JN, Ryan CP, McDade TW (2020) Evolutionary life history theory as an organizing framework for cohort studies: insights from the Cebu Longitudinal Health and Nutrition Survey. *Ann Hum Biol* 47:94–105
- Lettinga N, Jacquet PO, André JB, Baumard N, Chevallier C (2020) Environmental adversity is associated with lower investment in collective actions. *PLoS ONE* 15:e0236715
- Li KH, Meng XL, Raghunathan TE, Rubin DB (1991) Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica* 1:65–92
- MacCallum RC, Roznowski M, Necowitz LB (1992) Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychol Bull* 111:490–504
- McCullough ME, Pedersen EJ, Schroder JM, Tabak BA, Carver CS (2013) Harsh childhood environmental characteristics predict exploitation and retaliation in humans. *Proc Roy Soc B Biol Sci* 280:20122104
- McKay R, Whitehouse H (2015) Religion and morality. *Psychol Bull* 141:447–473
- Mell H, Baumard N, André JB (2019) Time is money. Waiting costs explain why selection favors steeper time discounting in deprived environments. *EcoEvoRxiv* <https://doi.org/10.32942/osf.io/7d56s>
- Moon JM, Krems JA, Cohen AB, Kenrick DT (2019) Is nothing sacred? religion, sex, and reproductive strategies. *Curr Dir Psychol Sci* 28:361–365
- Nieminen P, Lehtiniemi H, Vähäkangas K, Huusko A, Rautio A (2013) Standardised regression coefficient as an effect size index in summarising findings in epidemiological studies. *Epidemiol Biostat Public Health* 10:4
- Norenzayan A, Shariff AF, Gervais WM, Willard AK, McNamara RA, Slingerland E, Henrich J (2016) The cultural evolution of prosocial religions. *Behav Brain Sci* 39:e1
- Norris P, Inglehart R Religion and politics worl Sacred and secular: Religion and politics worldwide. Cambridge University Press.
- Obeyesekere G (2002) Imagining karma: Ethical transformation in Amerindian, Buddhist, and Greek rebirth (vol. 14). Univ of California Press.
- Pazhoohi F, Lang M, Xygalatas D, Grammer K (2017) Religious veiling as a mate-guarding strategy: effects of environmental pressures on cultural practices. *Evol Psychol Sci* 3:118–124
- Pearl J (2009) Causality: models, reasoning, and inference. Cambridge University Press, Cambridge, U.K
- Purzycki BG, Apicella C, Atkinson QD et al. (2016) Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature* 530:327–330
- Rosseel Y (2012) lavaan: an R package for structural equation modeling. *J Stat Softw* 48:1–36
- Ruck DJ, Bentley RA, Lawson DJ (2018) Religious change preceded economic change in the 20th century. *Sci Adv* 4:eaar8680
- Safra L, Tecu T, Lambert S, Sheskin M, Baumard N, Chevallier C (2016) Neighborhood deprivation negatively impacts children's prosocial behavior. *Front Psychol* 7:1760
- Santos FP, Santos FC, Pacheco JM (2016) Social norms of cooperation in small-scale societies. *PLoS Comp Biol* 12:e1004709
- Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66:507–514
- Segal A (2010) Life after death: A history of the afterlife in western religion. Image.
- Singh M (2018) The cultural evolution of shamanism. *Behav Brain Sci* 41:e66
- Truett KR (1993) Age differences in conservatism. *Pers Ind Diff* 14:405–411
- Turchin P (2011) Warfare and the evolution of social complexity: a multilevel selection approach. *Struct Dyn* 4:1–37
- van Buuren S, Groothuis-Oudshoorn K (2011) mice: Multivariate imputation by chained equations in R. *J Stat Softw* 45:1–67
- Weber M (1958) The protestant ethic and the spirit of capitalism. Scribner, New York [1904–1905]
- Weeden J, Cohen AB, Kenrick DT (2008) Religious attendance as reproductive support. *Evol Hum Behav* 29:327–334
- Weeden J, Kurzban R (2013) What predicts religiosity? A multinational analysis of reproductive and cooperative morals. *Evol Hum Behav* 34:440–445
- West SA, Griffin AS, Gardner A (2007) Evolutionary explanations for cooperation. *Curr Biol* 17:PR661–R672
- Whitehouse H, François P, Savage PE et al. (2019) Complex societies precede moralizing gods throughout world history. *Nature* 568:226–229
- Whittaker TA, Stapleton LM (2006) The Performance of cross-validation indices used to select among competing covariance structure models under multivariate nonnormality conditions. *Multivariate Behav Res* 41:295–335
- Wilson DS (2003) Darwin's Cathedral: Evolution, religion, and the nature of society. University of Chicago Press

Acknowledgements

We thank Jean-Baptiste André and Hugo Mercier for insightful conversations. This work was supported by the FrontCog funding under grant number ANR-17-EURE-0017. POJ was supported by ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL*. CF was supported by a graduate research fellowship from the Direction Générale de l'Armement (2015-60-0041).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information Supplementary information is available for this paper at <https://doi.org/10.1057/s41599-020-00691-9>.

Correspondence and requests for materials should be addressed to P.O.J. or N.B.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021