Explaining the Rise of Moralizing Religions
Supplementary Online Material

Methods

Seshat: Global History Databank

Seshat (http://seshatdatabank.info/) is a large database of information about global history from the Neolithic Revolution up to the Industrial Revolution (François et al., 2016; Turchin et al., 2015; Turchin et al., 2018). During the early stages of the project, we created an initial stratified sample of past societies by identifying 10 world regions distributed as widely as possible across the Earth's surface and within each of those regions designated three 'Natural Geographical Areas' (NGAs) with discrete ecological boundaries, on average about 10,000 km^2 in size, thus creating a sampling scheme of 30 such areas around the world (http://seshatdatabank.info/methods/world-sample-30/). The 30 regions and their selection rationale were published previously (Turchin et al., 2015) before the start of data collection. Our aim was to maximize variability in our global sample while minimizing historical relationships between cultures. We are in the process of adding further NGAs to the initial sample of 30 and at the time of writing Seshat contains 35 NGAs comprising 372 unique polities (see http://seshatdatabank.info/databrowser/).

Data on political systems (polities) that emerged and persisted in each of the NGAs are organized into a continuous time series. For the purposes of the present study, these are queried at 100-year intervals, going back as far into the history of that area as scholarly literature would allow (up to a maximum of roughly 10,000 years before present). In the case of NGAs containing clusters of very small-scale polities that share a similar culture but are not under a single system of jurisdictional control, we refer to these as 'quasi-polities' and code information on all of them generically, unless information is available that would allow us to differentiate between these polities.

All variables for which data have been gathered and entered into Seshat are derived from a Seshat Codebook that can be accessed and downloaded (http://seshatdatabank.info/methods/codebook). The Codebook was designed by, and is continually updated and extended in consultation with, a large network of professional historians, archaeologists, anthropologists and other specialists whom we refer to as 'Seshat experts'. Especially during the early phases of data entry, variables in the codebook were revised and improved through continuous discussions among Seshat research assistants, experts, and the Board. Most variables in Seshat require the data to take the form of a number or numerical range or they specify a feature that can be coded as absent, present or unknown (additionally coding items as 'inferred present' or 'inferred absent', where the evidence permits). All data are linked to scholarly sources, including peer-reviewed publications and personal communications from established authorities. A large subset of our dataset, including all variables used for this paper, can already be accessed at the project website http://seshatdatabank.info/databrowser/downloads.html (Turchin, Hoyer, et al., 2020).

While the Seshat project is constantly changing and evolving, many of our procedures and methods have become fairly standardized. Here, we redeploy certain descriptions about our methods and the construction of variables from other published work. This is done for parsimony and to reflect the interconnection between the various outputs from the Seshat team; together, these contribute to our collective effort to explore different hypotheses about the rise and fall of large-scale societies across the globe and human history.

Variables Used in the Analysis

Moralizing Supernatural Punishment (MSP)

Ten variables pertaining to moralizing supernatural punishment/reward were coded, as follows:

Primary: MSP is described as 'primary' when the principal concerns of supernatural agents or forces pertain to cooperation in human affairs. It is coded as absent when the primary concern is the behavior of humans towards the supernatural realm, e.g. by discharging ritual obligations. Importantly, codings of 'primary' were applied not only when moralizing religion took the form of a punitive agent but also when beliefs in non-agentic forms of supernatural enforcement were present, including karmic principles emphasizing incentives to behave morally as well as punishments for transgressions.

Certain: This variable reflects the predictability of supernatural punishment for transgression or reward for ethical behavior. A code of absence here could result from a variety of characteristics of supernatural agents: if they are fickle or capricious, if they can be bought off or tricked, or, alternatively, if they are not independently concerned about human morality and need to be persuaded or induced to punish transgressions.

Broad: This reflects how many aspects of morality deities care about and enforce. It is coded as absent when moralizing supernatural punishment/reward pertains to only very narrowly circumscribed domains, for example, kin-based moral precepts punishing incest or rewarding hospitality rather than enforcing moral norms across a broad range of social situations.

Targeted: This reflects whether punishment and rewards are targeted specifically at culpable individuals. It is coded as absent when the whole group is punished rather than just the individual transgressor.

Ruler: This reflects whether supernatural forces or agents punish/reward rulers for their antisocial/prosocial behavior. It can be absent where such punishment is present generally, but rulers remain exempt.

Elites: This reflects whether elites of the polity subscribe to a religion with moralizing elements. In some cases, only a vocal segment of the elites advocated a particular moralizing religion (for example, early Buddhists, some Christians, Confucians) but not entire elite populations.

Commoners: This reflects the extent to which beliefs in MSP are adopted by the masses. A typical situation in which this variable is coded absent is when the state religion professed by rulers and elites, and endorsing beliefs in supernatural punishments and rewards, is different from the popular religion which lacks or professes only much weaker beliefs in supernatural enforcement. On the other hand, this variable might be coded as present, even while the Elites variable is coded absent, for example when popular religion emphasizes supernatural enforcement, but the religion of rulers and elites does not. Assessing the beliefs of commoners is methodologically challenging. Depending on the period and polity, different types of evidence may be used to determine whether belief in supernatural punishments and/or rewards was widely distributed. In the prehistoric periods of Latium, for example, we used linguistic evidence (comparisons of oath formulas across a broad range of Indo-European languages) to code "inferred present." For the historical period, in addition to linguistic evidence, we used written sources such as popular comic plays with moralizing sentiment and expectations of MSP. In polities where a moralizing religion (i.e. one with primary concern for interpersonal cooperation) has been installed for a long time, we generally code MSP "present" for commoners, making due allowance for transitional periods.

AfterLife: Moralizing enforcement in afterlife, reflecting whether punishment is delayed until after the death of the transgressor.

ThisLife: Moralizing enforcement in this life. Reflects whether punishment occurs during transgressor's lifetime. It is possible to code both this variable and AfterLife as present, if punishment can occur both in this life and in afterlife.

Agency: Moralizing enforcement is agentic. Reflects whether punishment/reward is administered by a supernatural agent, such as a deity or spirit (as opposed to being administered by an impersonal supernatural force, such as karma).

Our main measure of moralizing supernatural punishment (MSP) is based on the first seven MSP characteristics in the list above (**Primary** through **Commoners**). If all characteristics were present, the aggregated moralizing religion variable was set to 1 (the maximum). Each code of absent reduced the maximum by half; that is, the overall score was multiplied by 0.5. The minimum of the aggregated measure, thus, is $0.5^7 \approx 0.008$. Unknowns were treated as missing data and are dropped from the analysis.

This procedure assumes multiplicative effects. We also reran all analyses with an alternative, additive aggregation scheme (equating present with 1, absent with 0, absent/present with 0.5, and adding together these numerical scores).

The resulting MSP measure (whether multiplicative or additive) is a categorical variable with 15 levels (due to "half-tones" introduced by transitional periods absent/present). It is used as the response (dependent) variable in dynamic regression analyses.

The last three variables (**AfterLife** through **Agency**) were used to explore whether the immediacy of punishment (in this life, or the afterlife) and the mechanism of punishment (by a supernatural agent or supernatural force) affects our results. To do this we constructed three additional measures that reflected only moralizing punishment/reward in the afterlife, only that in this life, and only that administered by supernatural agents. Thus, MSP_{after}, relying on punishment in the afterlife, was calculated by setting MSP to zero if **AfterLife** = absent. The other two measures, MSP_{this} and MSP_{agen}, were constructed analogously by setting MSP to zero if **ThisLife** or **Agency** were coded as absent.

Predictor Variables

We constructed several predictor variables theorized to interact with moralizing supernatural punishment/reward, as outlined in the main text. These include:

Sociopolitical complexity: The relationship between moralizing supernatural punishment/reward and the evolution of sociopolitical complexity has been a focus of much recent scholarship, as detailed in the main text. We measure socio-political complexity by a composite of 51 separate variables, following (Turchin et al., 2018). These variables were first aggregated into eight composite categories ("Complexity Characteristics") capturing different dimensions of complexity: polity population size, capital population size, polity territory size, hierarchy, infrastructure, government, information systems, and monetary instruments (Turchin et al. 2018 provides further details on each of the 51 variables). Statistical analysis of the Complexity Characteristics found that they all closely correlated and that the first Principal Component captures more than three-quarters of variance in the data (Turchin et al., 2018). We use this first principal component here as our measure of socio-political complexity, **SPC1**. In order to make SPC1 easily interpretable, we scale it in such a way that it corresponds to $\log_{10}$(Polity Population). In other words, polities with SPC1 = 3 have, on average, populations of 1000, and SPC1 = 6 corresponds to polities with populations of 1,000,000.

Warfare and interpolity competition: We proxy the intensity of interpolity competition by measuring military technology in two ways. The first proxy aggregates 46 variables measuring the realized sophistication and variety of military technologies in Seshat polities, **MilTech**. These variables code for the presence or absence of various types of technology in six composite categories: handheld weapons, armor, projectiles, and defensive structures, as well as the use of metals for making weapons and armor, and of transport animals used for military logistics. We describe these as “realized” technologies, as our coding approach assigns 1 when there is evidence that a particular weapon, projectile, etc. was used by the coded society and 0 when such evidence is absent. The reason for this “strong evidence” scheme is that our focus is not on whether a technology was known, but whether it was used. A large variety of sophisticated means of attack and defense, thus, serves as a quantitative proxy for the intensity of warfare in the environment of the polity. The MilTech measure used here is the sum of the six composite categories, which are, in turn, aggregated using the above scheme. Thus, the total range over which MilTech can vary is 0–46. Details on the 46 variables and methods of aggregation are in (Turchin, Korotayev, et al., 2020).

The second proxy we use to measure interpolity competition is the spread of cavalry, measured as the presence or absence of horse-mounted warfare. We code this through a single binary variable, **Cavalry**. Data on this variable is taken from (Turchin et al., 2016). Previous research suggested that horse-mounted warfare in particular is an important predictor influencing the evolution of the social scale and complexity of polities, beyond the influence of military technologies generally (Turchin, 2009; Turchin et al., 2013). The Cavalry variable differs from the ‘Horse’ variable included in the MilTech measure as Horse codes the use of horses in military activity including logistics (such as draft or pack animals), whereas Cavalry measures the adoption of a package of technological and tactical features employed in mounted warfare.

Agriculture: The Seshat project has developed a sophisticated approach to estimating how the productivity of agriculture has evolved in each of the Seshat Natural Geographic Areas (NGAs) on which the Seshat Sample of past polities is based (Turchin et al., 2021). The approach that we used to obtain these quantiative estimates combined the influences of production technologies (and how they change with time), climate change, and effects of artificial selection into a Relative Yield Coefficient, indicating how agricultural productivity changed over time in each NGA between the Neolithic and the 20th century. We then use estimates of historical yield in each NGA to translate the Relative Yield Coefficient into an Estimated Yield (tons per hectare per year) trajectory, which we use here as our measure of productivity, **Agri**. We tested the proposed methodology with independent data and concluded that while more work is needed to refine this approach, it provides reasonable approximation of agricultural productivities for the societies in the Seshat sample (Turchin et al., 2021).

Pastoralism: We use the data recently published by the ArchaeoGlobe project (Stephens et al., 2019). This project synthesized the knowledge of c.250 archaeologists who have coded 146 world regions (“AG regions”) for the presence of pastoralism (as well as foraging, extensive and intensive agriculture, and urbanization, but our focus is on pastoralism) at 10 time intervals stretching from 10k BP (8,000 BCE) to 1850. ArchaeoGlobe experts coded each AG region for each time step for pastoralism, **Pastor**, using a categorical scale with four levels, which we translated into a numerical range. These levels and associated numbers are 0: none (no evidence that any land in the region was used for pastoralism), 1: minimal (pastoralism was present, but less than 1% of land in the region was used for it), 2: common (between 1% and 20% of land was used for pastoralism), and 3: widespread (greater than 20% of land was used for pastoralism).

Environment: We used the environmental data compiled by Botero et al. (Botero et al., 2014). Principal Component analysis of the environmental characteristics coded by that project—average rainfall, temperature variability, abundance of available flora and fauna species—produced two principal components, which we use here as our two measures of environmental predictability and resource scarcity: **EnvPC1** and **EnvPC2** (see *Results* below for the PC analysis).

Data gathering and collation

MSP and moralizing religion: Our data gathering strategy followed a transparent and rigorous process taking place over several years and involving project experts, research assistants, and a Data Review Board (DRB) (see http://seshatdatabank.info/methods/). The latter comprises the senior team responsible for data management on a given paper. For the present paper, the DRB included three historians (DH, PF, and JL), an anthropologist (HW), and a complexity scientist (PT). The process of data collection for MSP variables typically involved matching each of the fully trained research assistants with one or more experts (recognized authorities on the polity in question, typically holding a relevant doctorate and occupying a faculty position in a university). Initial input by the experts focused on providing help with assembling initial reading lists or, where necessary, advice on how to interpret some of the key historiographical debates. Research assistants gathered the information necessary to put forward a provisional coding recommendation, together with a condensed overview of the data used to buttress that coding, highlighting any areas of uncertainty. These codes are thus based on scholarly sources and are fully referenced. In addition to the codes of 'absence', 'presence', and 'unknown', a coding of 'inferred' absence (or presence) was used when direct evidence for a particular variable was sparse or lacking but indirect evidence made clear that it was more likely to have been absent (or present) than not. This approach avoids a situation in which researchers inaccurately coded the trait 'unknown' when in fact what was known was more than nothing. In addition, variables could be coded as first absent but then present during transitional periods or could be coded in multiple ways simultaneously where experts disagreed, thus providing grounds for more than one coding outcome. Research assistants then conducted consistency checks. Coding recommendations and the data provisionally used to buttress them were then presented to experts for further review, often in multiple iterations. Where research assistants found no information on a particular variable, they assigned a temporary code of 'suspected unknown', which was later converted to 'unknown' after being confirmed by an expert.

When Seshat experts pointed out disagreements in the literature or disagreed among themselves on a particular coding, we kept a record of this so that multiple analyses could be run taking into account contrasting interpretations. Finally, the DRB reviewed the resulting coding recommendations and supporting data. At this stage the DRB could approve codes as ready for analysis or request further review, where appropriate, involving additional experts to address remaining points of uncertainty. The DRB was also responsible for ensuring at this point that coding conventions were consistently applied across NGAs. Only when the DRB was satisfied that the rationales for coding decisions and the associated buttressing statements were transparently and compellingly articulated, following a set of agreed coding conventions, were the data 'frozen' and converted into the correct syntax for the analysis. As such, final responsibility for coding decisions relating to data frozen for publication was assumed by the DRB rather than being outsourced to contributing experts.

Predictor variables: the procedures for collecting these data are detailed in special publications: socio-political complexity (Turchin et al., 2018), MilTech (Turchin, Korotayev, et al., 2020), Cavalry (Turchin et al., 2016), agriculture (Turchin et al., 2021), pastoralism (Stephens et al., 2019), and environmental variables (Botero et al., 2014).

Results

Environmental Variables

Principal Component Analysis indicated that the first two principal components together capture over three-quarters of variation in environmental data (Table S1)

Table S1. Results of PCA:

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	1.8635	1.3583	1.0191	0.57225	0.39677	0.3051	0.25694
Proportion of Variance	0.4961	0.2636	0.1484	0.04678	0.02249	0.0133	0.00943
Cumulative Proportion	0.4961	0.7596	0.9080	0.95478	0.97727	0.9906	1.00000

Rotation (n x k) = (7 x 7):

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
MeanP1950	0.45	-0.32	0.268	-0.152	0.093	0.1116	-0.765
VarP1950	0.32	-0.29	0.674	-0.082	0.126	0.0574	0.580
Pp1950	0.31	-0.47	-0.364	0.497	-0.412	0.3319	0.147
MeanT1950	0.26	0.62	0.054	-0.063	-0.039	0.7335	0.030
VarT1950	-0.40	-0.37	-0.069	-0.673	-0.299	0.3866	0.048
Pt1950	0.47	0.24	-0.081	-0.371	-0.621	-0.4318	0.082
MeanN	0.39	-0.12	-0.572	-0.357	0.574	-0.0079	0.217

In our analysis we used these two components, EnvPC1 and EnvPC2. Mapping Seshat NGAs by these environmental principal components suggest that EnvPC1 captures the overall suitability of the NGA for agriculture, while EnvPC2 reflects the cline from cold to hot and dry environments (Figure S1).

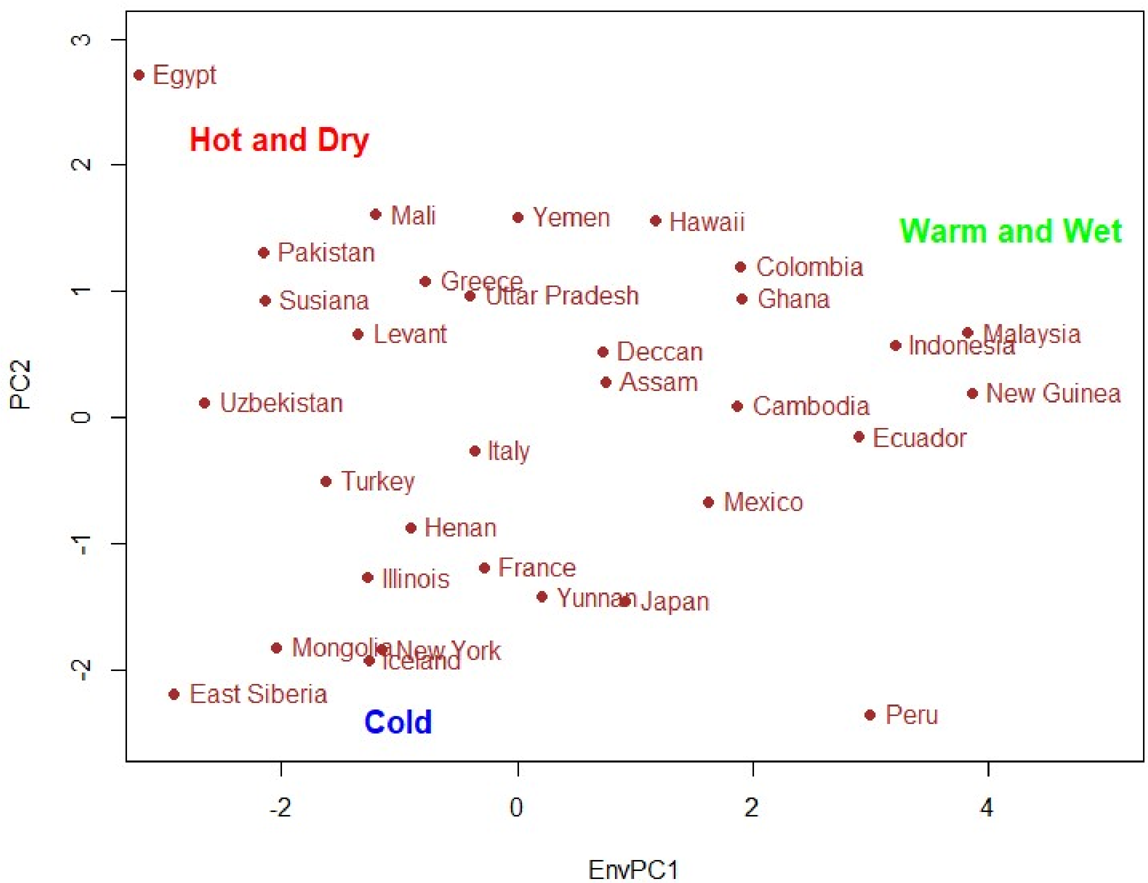

Figure S1. Environmental variables (the first two principal components summarizing mean, variation, and predictability in temperature and precipitation) differ from other variables because they only depend on spatial location (NGA) and do not change with time.

Descriptive Statistics

The distributions and correlations between MSP and potential predictor variables are shown in Figure S2.

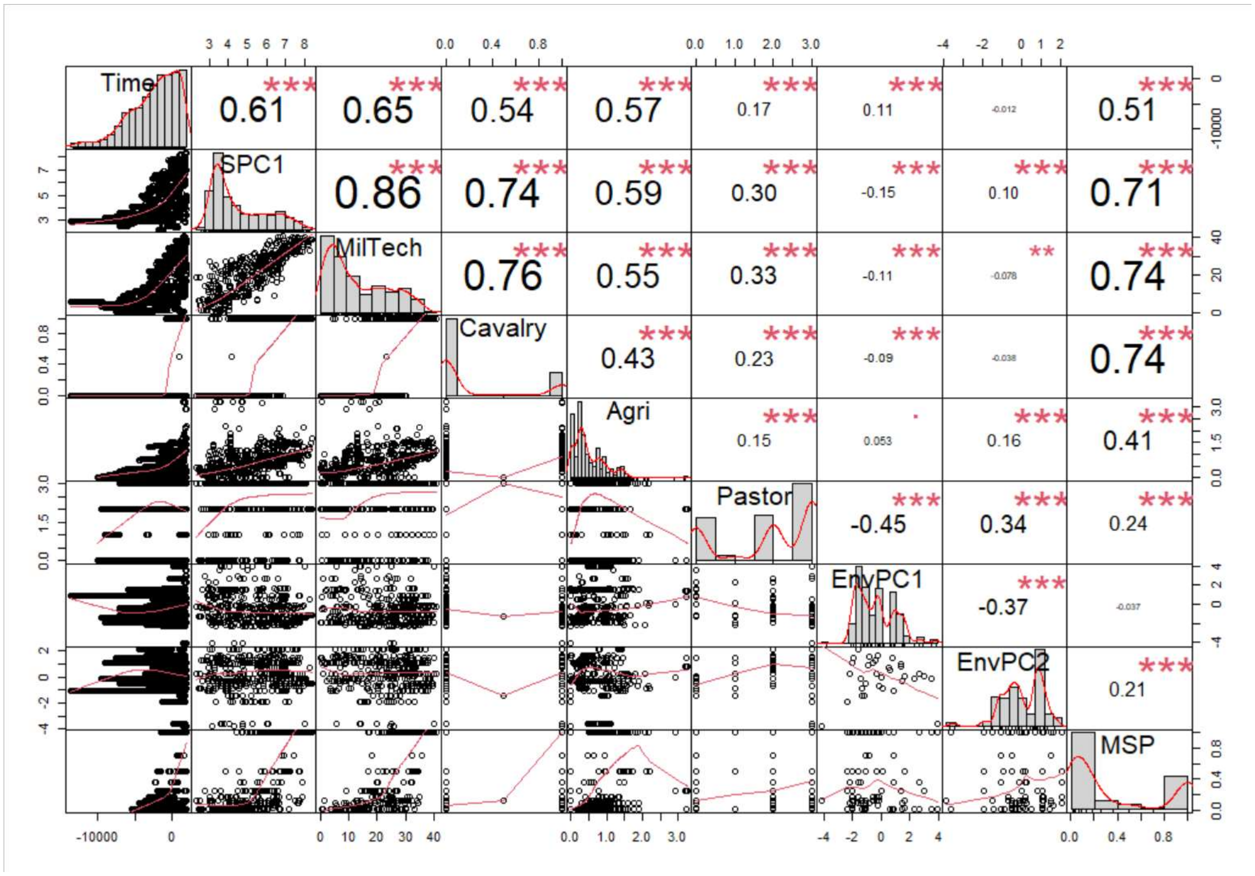

Figure S2. Summary statistics. The diagonal cells plot frequency distribution for each variable. Below the diagonal are pairwise scatter plots. Above the diagonal are the values of pairwise Pearson correlation coefficients. This plot was produced by the chart.Correlation function from PerformanceAnalytics library.

Dynamic Regression: MSP as the response

First, we investigate which combination of the potential predictors best explains the variation in MSP. Because the correlation plot of MSP against Agri suggested presence of nonlinearity (Figure S2), we include a quadratic term for this potential predictor. Exhaustive regressions with these terms yield the following table of models ranked by their AIC (Table S2) and full regression results for the best model (Table S3).

Table S2. Regression results: MSP. This table shows the best models with delAIC < 2 (out of >500 total configurations). Each row provides the t-statistic associated with the predictor terms (an empty cell indicates that this particular term was not used in regression).

MSP	MSP.sq	SPC1	MilTech	Cavalry	Agri	Agri.sq	Pastor	EnvPC1	EnvPC2	delAIC
10.835	-4.212		2.366	4.814	2.804	-2.115	2.318	2.196	2.582	0.000
10.905	-3.969	-1.349	2.725	5.002	3.063	-2.354	2.383	2.152	2.862	0.152

Table S3. Regression results for the best model (by AIC).

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.399e-15  1.094e-02   0.000   1.0000    
MSP          6.130e-01  5.657e-02  10.835  < 2e-16 ***
MSP.sq      -2.002e-01  4.751e-02  -4.212 2.89e-05 ***
MilTech      5.216e-02  2.204e-02   2.366   0.0183 *  
Cavalry      9.465e-02  1.966e-02   4.814 1.84e-06 ***
Agri         9.495e-02  3.386e-02   2.804   0.0052 ** 
Agri.sq     -6.114e-02  2.891e-02  -2.115   0.0348 *  
Pastor       3.124e-02  1.348e-02   2.318   0.0208 *  
EnvPC1       2.978e-02  1.356e-02   2.196   0.0285 *  
EnvPC2       3.701e-02  1.433e-02   2.582   0.0100 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2802 on 646 degrees of freedom
Multiple R-squared:  0.9226,  Adjusted R-squared:  0.9215 
F-statistic: 855.4 on 9 and 646 DF,  p-value: < 2.2e-16
```

Checking for the effects of autoregressive terms indicates that they do not add to the explanatory value of the best model. Most importantly, the addition of autoregressive terms does not change the main result. Similarly, we run a regression including time as a covariate, as a check for omitted variables (Turchin, 2018), and discover no improvement. Including NGA as a fixed effect, as expected, makes the environmental variables (EnvPC1 and EnvPC2) not significant, but doesn't affect MilTech, Cavalry, and Agri terms.

Performing regression diagnostics we find that the distribution of residuals clearly doesn't conform to the Normal distribution (Figure S3)

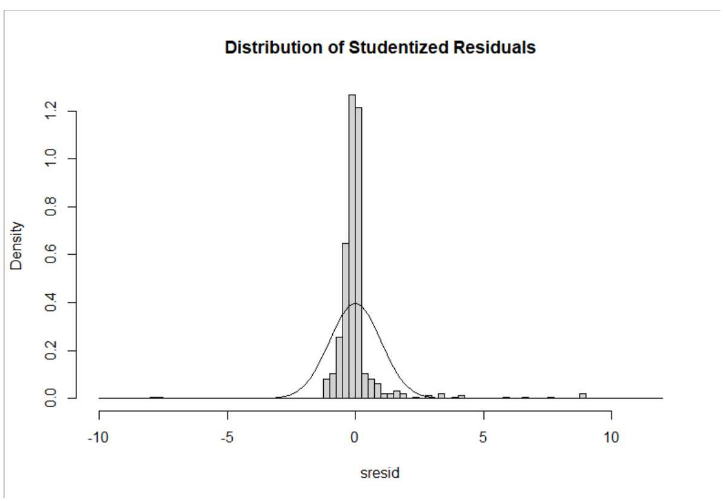

Figure S3. Distribution of studentized residuals of the best-fitting model.

For this reason, we use nonparametric bootstrap to approximate the 95% confidence intervals for the regression estimates (Table S4).

Table S4. Nonparametric bootstrap results.

	Lower	Upper	Prop > 0*
MSP(t+1)	-0.024	0.027	0.501
MSP	0.451	0.740	0.000
MSP.sq	-0.308	-0.091	0.000
MilTech	0.007	0.111	0.016
Cavalry	0.037	0.156	0.000
Agri	0.005	0.198	0.019
Agri.sq	-0.136	0.016	0.044
Pastor	0.005	0.069	0.008
EnvPC1	-0.005	0.077	0.052
EnvPC2	-0.002	0.086	0.033

* *Or Prop < 0 for negative terms (MSP.sq and Agri.sq)*

Bootstrap results indicate that the AIC-based results are somewhat liberal (tend to inflate the statistical significance of regression terms). Thus, the 95% confidence intervals for Agri.sq, EnvPC1, and EnvPC2 overlap 0. This means that the evidence for the effect on MSP by such environmental and economic factors is substantially weaker, compared to the effects of war intensity proxies (MilTech and Cavalry).

Overall, this investigation suggests that the main effect on the evolution of MSP, as measured by the absolute value of the standardized coefficient, is by Cavalry, followed by Agri and MilTech, and then the rest of the terms. Once these terms are included, there is no evidence that SPC1 has a causal effect. Even when it is included in the models (of worse AIC than the best one), its effect is weak (small *t*) and statistically insignificant.

Additional checks support this general conclusion. First, using an alternative measure of MSP that sums together the MSP components (instead of the main multiplicative measure) yields identical results, because log-transforming multiplicative MSP is linearly related to additive MSP.

Next, we check how *not* deleting data that were not yet approved by the Data Review Board (due to insufficient expert feedback) (Table S5).

Table S5. Regression results for the full dataset (including data not approved by the DRB)

MSP	MSP.sq	SPC1	MilTech	Cavalry	Agri	Agri.sq	Pastor	EnvPC1	EnvPC2	delAIC
10.382	-4.542		2.728	5.356	2.685	-2.066	1.914	2.453	2.739	
10.447	-4.320	-1.293	3.013	5.504	2.929	-2.292	1.950	2.418	2.995	0.302
10.347	-4.843		2.920	5.242	2.410	-1.955		2.120	2.741	1.707

Finally, we check how excluding inferred absence (A*) codes affects our results. The most dramatic effect is the decrease in sample size, from 656 data points in the main data set, to 350 data in the set where A* are omitted. As expected, such severe reduction of the data decreases statistical power of the analysis (ability to detect true effects). Nevertheless, the strongest predictors, Cavalry and Agri, continue to be supported in the analyses with restricted data. Additionally, EnvPC2 is selected in the majority of models with delAIC < 2. The negative result (no effect of SPC1 on MSP) is also supported.

Table S6. Regression results for the data set in which all inferred absence data are excluded.

MSP	MSP.sq	SPC1	MilTech	Cavalry	Agri	Agri.sq	Pastor	EnvPC1	EnvPC2	delAIC
29.950				4.623		2.928			1.572	
29.825				4.677	-1.344	2.240			2.001	0.166
32.946				4.380		2.931				0.499
29.586		-1.012		4.680		3.042			1.600	0.960
29.768				4.687	-1.521	2.364		0.887	2.162	1.365
32.361		-0.966		4.415		3.039				1.554
15.814	0.546			4.369		2.920			1.567	1.697
29.905				4.620		2.808		0.526	1.637	1.718
29.222			-0.452	4.113		2.940			1.485	1.792
29.564		-0.537		4.541	-1.034	1.945			1.894	1.872
28.978				4.603		2.856	0.313		1.529	1.901
15.867	0.474			4.434	-1.315	2.209			1.984	1.936

Effect of MSP on SPC1

To investigate whether there is a feedback effect of MSP on SPC1, we repeat the analysis but with SPC1 as the response variable and adding MSP to the set of potential predictors. With this set of possible predictors, there are only two models with delAIC < 2. We observe that while MSP is selected for the model with lowest AIC, the coefficient associated with it is negative and statistically non-significant (Tables S7 and S8).

Table S7. Regression results for SPC1 as the response variable: models with delAIC < 2.

SPC1	SPC1.sq	MilTech	Cavalry	Agri	MSP	Lag2	delAIC
10.146	-3.916	2.503	3.509	3.240	-1.847	4.428	
10.381	-4.096	1.941	3.234	2.766		4.274	1.449

Table S8. Regression results for SPC1 as the response variable (the best model by AIC).

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.005879   0.012585  -0.467 0.640565
SPC1         1.079742   0.106424  10.146  < 2e-16 ***
SPC1.sq     -0.395890   0.101099  -3.916 0.000101 ***
MilTech      0.064552   0.025787   2.503 0.012573 *
Cavalry      0.072752   0.020734   3.509 0.000484 ***
Agri         0.048909   0.015095   3.240 0.001262 **
MSP         -0.038120   0.020634  -1.847 0.065187 .
Lag2         0.172819   0.039028   4.428 1.13e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.308 on 592 degrees of freedom
```

```
Multiple R-squared:  0.9057,  Adjusted R-squared:  0.9046
F-statistic: 811.9 on 7 and 592 DF,  p-value: < 2.2e-16
```

One difference between these results and the MSP analysis is that SPC1 is strongly affected by Lag2. In other words, it's an autoregressive process of order 2, unlike MSP, which is AR(1). Of the potential predictors, the strongest effects are by Cavalry and MilTech, followed by Agri. These results yield strong evidence against the effect of MSP on SPC1, because the estimated coefficient associated with the MSP term is negative and is not statistically significant at $P < 0.05$ level.

Effect of Alternative MSP Specifications

We now test whether alternative ways of quantifying MSP affect our conclusion that there is no causal arrow from MSP to SPC1. First, we use Minimal MSP (minMSP, see *Methods* in the main article) as a possible predictor for SPC1. While the numbers change, the qualitative result is the same as with full MSP. Despite minMSP preceding MSP by 1000 years on average, there is no support for the hypothesis that minMSP has a positive effect on SPC1 (Table S9a). We also find that different versions of MSP (ThisLife, AfterLife, and Agency, see *Methods*) similarly do not have statistically detectable effects on SPC1 (Table S9b,c,d).

Table S9. Effect of alternative MSP specifications.

(a) Minimal MSP

SPC1	**SPC1.sq**	**MilTech**	**Cavalry**	**Agri**	**minMSP**	**Lag2**	**delAIC**
10.381	-4.096	1.941	3.234	2.766		4.274	
10.355	-4.083	2.078	3.200	2.857	-0.743	4.279	1.440
10.208	-3.747		4.472	2.810		4.352	1.800

(b) MSP_{this}

SPC1	**SPC1.sq**	**MilTech**	**Cavalry**	**Agri**	**MSP_this**	**Lag2**	**delAIC**
9.640	-3.896	2.367	2.736	2.135		4.374	
9.514	-3.856	2.180	2.729	2.102	0.091	4.345	1.992

(c) MSP_{after}

SPC1	**SPC1.sq**	**MilTech**	**Cavalry**	**Agri**	**MSP_after**	**Lag2**	**delAIC**
8.348	-3.067	2.557	2.861	3.280	-1.684	4.166	
8.440	-3.217	2.506	2.357	2.857		4.006	0.871

(d) MSP_{agen}

SPC1	**SPC1.sq**	**MilTech**	**Cavalry**	**Agri**	**MSP_agen**	**Lag2**	**delAIC**
9.959	-4.153	2.200	3.077	2.494		4.169	
9.920	-4.144	2.229	2.988	2.533	-0.472	4.189	1.774

Evolutionary Drivers of Sociopolitical Complexity

Finally, having shown that evidence against the positive effect of MSP on SPC1 is strong and consistent, we drop it from the analysis. This enables us to utilize the complete dataset (not needing to remove rows with missing MSP values), which results in higher statistical power to detect possible causal effects by predictor variables. Apart from autoregressive terms (SPC1, SPC1.sq, Phylogeny, and Lag2) only three predictors are selected: Cavalry, MilTech, and Agri (in order of the magnitude of the standardized regression coefficient). Statistical evidence for all three is very strong, as indicated by low *P*-values (Tables S10 and S11).

Table S10. The results of the exhaustive search with SPC1 as response.

SPC1	SPC1.sq	MilTech	Cavalry	Agri	Agri.sq	Pastor	EnvPC1	EnvPC2	Lag2	delAIC
18.002	-7.127	4.012	5.803	4.236					4.539	
18.036	-7.158	4.003	5.731	4.001			1.279		4.620	0.354
17.971	-7.124	3.844	5.855	2.745	-0.796				4.486	1.362
18.023	-7.184	4.077	5.770	3.575			1.558	0.948	4.578	1.450
18.006	-7.156	3.826	5.793	2.697	-0.851		1.314		4.568	1.624
17.736	-6.943	3.787	5.717	3.952		0.841	1.522		4.642	1.642
17.983	-7.130	3.892	5.810	4.083				0.326	4.505	1.893
17.799	-7.022	3.935	5.802	4.237		0.163			4.536	1.973

Table S11. Full results for the best supported model are:

	Estimate	SE	t-value	Pr(>|t|)
(Intercept)	-0.005	0.007	-0.711	0.477074
SPC1	1.137	0.063	18.014	0.000000
SPC1.sq	-0.414	0.058	-7.145	0.000000
MilTech	0.051	0.014	3.558	0.000388
Cavalry	0.060	0.012	5.100	0.000000
Agri	0.036	0.009	3.999	0.000067
Phylogeny	0.023	0.010	2.308	0.021180
Lag2	0.118	0.026	4.562	0.000006

A final note is that the full analysis of the evolutionary drivers of SPC, which differentiates its separate components (Social Scale, Hierarchy, and Sophistication of Governance Institutions) and tests 17 hypotheses derived from the current theoretical corpus is reported in (Turchin et al. in prep).

References

Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C., & Gray, R. D. (2014). The ecology of religious beliefs. *PNAS*. www.pnas.org/cgi/doi/10.1073/pnas.1408701111

François, P., Manning, J. G., Whitehouse, H., Brennan, R., Currie, T., Feeney, K., & Turchin, P. (2016). A Macroscope for Global History: Seshat Global History Databank, a methodological overview.

Digital Humanities Quarterly, *10*(4). http://www.digitalhumanities.org/dhq/vol/10/4/000272/000272.html

Stephens, L., Fuller, D., Boivin, N., Rick, T., Gauthier, N., Kay, A., Marwick, B., Geralda, C., Armstrong, D., Barton, C. M., Denham, T., Douglass, K., Driver, J., Janz, L., Roberts, P., Rogers, J. D., Thakar, H., Altaweel, M., Johnson, A. L., Sampietro Vattuone, M. M., Aldenderfer, M., Archila, S., Artioli, G., Bale, M. T., Beach, T., Borrell, F., Braje, T., Buckland, P. I., Jiménez Cano, N. G., Capriles, J. M., Diez Castillo, A., Çilingiroğlu, Ç., Negus Cleary, M., Conolly, J., Coutros, P. R., Covey, R. A., Cremaschi, M., Crowther, A., Der, L., di Lernia, S., Doershuk, J. F., Doolittle, W. E., Edwards, K. J., Erlandson, J. M., Evans, D., Fairbairn, A., Faulkner, P., Feinman, G., Fernandes, R., Fitzpatrick, S. M., Fyfe, R., Garcea, E., Goldstein, S., Goodman, R. C., Dalpoim Guedes, J., Herrmann, J., Hiscock, P., Hommel, P., Horsburgh, K. A., Hritz, C., Ives, J. W., Junno, A., Kahn, J. G., Kaufman, B., Kearns, C., Kidder, T. R., Lanoë, F., Lawrence, D., Lee, G.-A., Levin, M. J., Lindskoug, H. B., López-Sáez, J. A., Macrae, S., Marchant, R., Marston, J. M., McClure, S., McCoy, M. D., Miller, A. V., Morrison, M., Motuzaite Matuzeviciute, G., Müller, J., Nayak, A., Noerwidi, S., Peres, T. M., Peterson, C. E., Proctor, L., Randall, A. R., Renette, S., Robbins Schug, G., Ryzewski, K., Saini, R., Scheinsohn, V., Schmidt, P., Sebillaud, P., Seitsonen, O., Simpson, I. A., Sołtysiak, A., Speakman, R. J., Spengler, R. N., Steffen, M. L., Storozum, M. J., Strickland, K. M., Thompson, J., Thurston, T. L., Ulm, S., Ustunkaya, M. C., Welker, M. H., West, C., Williams, P. R., Wright, D. K., Wright, N., Zahir, M., Zerboni, A., Beaudoin, E., Munevar Garcia, S., Powell, J., Thornton, A., Kaplan, J. O., Gaillard, M.-J., Klein Goldewijk, K., & Ellis, E. (2019). Archaeological assessment reveals Earth's early transformation through land use. *Science*, *365*(6456), 897-902. https://doi.org/10.1126/science.aax1192

Turchin, P. (2009). A Theory for Formation of Large Empires. *Journal of Global History*, *4*, 191-207.

Turchin, P. (2018). Fitting Dynamic Regression Models to Seshat Data. *Cliodynamics*, *9*, 25-58.

Turchin, P., Brennan, R., Currie, T., Feeney, K., Francois, P., Hoyer, D., Manning, J., Marciniak, A., Mullins, D., Palmisano, A., Peregrine, P., Turner, E. A. L., & Whitehouse, H. (2015). Seshat: The Global History Databank. *Cliodynamics*, *6*, 77-107.

Turchin, P., Currie, T. E., Collins, C., Levine, J., Oyebamiji, O., Edwards, N. R., Holden, P. B., Hoyer, D., Feeney, K., Francois, P., & Whitehouse, H. (2021). An integrative approach to estimating productivity in past societies using Seshat: Global History Databank. *Holocene*, *31*(6), 1055-1065. https://doi.org/10.1177/0959683621994644

Turchin, P., Currie, T. E., & Turner, E. A. L. (2016). Mapping the Spread of Mounted Warfare. *Cliodynamics*, *7*, 217-227. https://doi.org/https://doi.org/10.21237/C7clio7233509

Turchin, P., Currie, T. E., Turner, E. A. L., & Gavrilets, S. (2013). War, Space, and the Evolution of Old World Complex Societies. *PNAS*, *110*, 16384–16389.

Turchin, P., Currie, T. E., Whitehouse, H., Francois, P., Feeney, K., Mullins, D., Hoyer, D., Collins, C., Grohmann, S., Savage, P., Mendel-Gleason, G., Turner, E., Dupeyron, A., Cioni, E., Reddish, J., Levine, J., Jordan, G., Brandl, E., Williams, A., Cesaretti, R., Krueger, M., Ceccarelli, A., Figliulo-Rosswurm, J., Tuan, P.-J., Peregrine, P., Marciniak, A., Preiser-Kapeller, J., Kradin, N., Korotayev, A., Palmisano, A., Baker, D., Bidmead, J., Bol, P., Christian, D., Cook, C., Covey, A., Feinman, G., Juliusson, A. D., Kristinsson, A., Miksic, J., Mostern, R., Petrie, C., Rudiak-Gould, P., Ter Haar, B., Wallace, V., Mair, V., Xie, L., Baines, J., Bridges, E., Manning, J., Lockhart, B., Bogaard, A., & Spencer, C. (2018). Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proc Natl Acad Sci U S A*, *115*(2), e144-e151. https://doi.org/10.1073/pnas.1708800115

Turchin, P., Hoyer, D., Bennett, J., Basava, K., Cioni, E., Francois, P., Holder, S., Levine, J., Nugent, S., Reddish, J., Thorpe, C., Wiltshire, S., & Whitehouse, H. (2020). The Equinox2020 Seshat Data Release. *Cliodynamics*, *11*(1), 41-50. https://doi.org/https://doi.org/10.21237/C7clio11148620

Turchin, P., Korotayev, A., Kradin, N., Nefedov, S., Feinman, G., Bennett, J. S., Hoyer, D., Francois, P., & Whitehouse, H. (2020). Rise of the War Machines: Charting the Evolution of Military Technologies from the Neolithic to the Industrial Revolution. *SocArXiv Preprint*.