Treatment of missing data determines conclusions regarding moralizing gods

Bret Beheim*, Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology

Quentin Atkinson, School of Psychology, University of Auckland

Joseph Bulbulia, School of Humanities, University of Auckland

Will Gervais, Department of Psychology, University of Kentucky

Russell D. Gray, Max Planck Institute for the Science of Human History, School of Psychology, University of Auckland

Joseph Henrich, Department of Human Evolutionary Biology, Harvard University

Martin Lang, LEVYNA: Laboratory for the Experimental Research of Religion, Masaryk University

M. Willis Monroe, Department of Asian Studies, University of British Columbia

Michael Muthukrishna, Department of Psychological and Behavioural Science, London School of Economics

Ara Norenzayan, Department of Psychology, University of British Columbia

Benjamin Grant Purzycki, Department of the Study of Religion, Aarhus University

Azim Shariff, Department of Psychology, University of British Columbia

Edward Slingerland, Department of Asian Studies, University of British Columbia

Rachel Spicer, Department of Psychological and Behavioural Science, London School of Economics

Aiyana K. Willard, Centre for Culture and Evolution, Brunel University London

* corresponding author; bret_beheim@eva.mpg.de

Whitehouse, *et al*.[1] have recently used the Seshat archaeo-historical databank[2] to argue that beliefs in moralizing gods appear in world history only after the formation of complex "megasocieties" of around one million people. Inspection of the authors' data, however, shows that 61% of Seshat data points on moralizing gods are missing values, mostly from smaller populations below one million people, and during the analysis the authors re-coded these data points to signify the *absence* of moralizing gods beliefs. When we confine the analysis only to the extant data or use various standard imputation methods, the reported finding is reversed: moralizing gods precede increases in social complexity. We suggest that the reported "megasociety threshold" for the emergence of moralizing gods is thus solely a consequence of the decision to re-code nearly two-thirds of Seshat data from unknown values to known absences of moralizing gods.

Table 1 illustrates the Seshat data for the 12 key world regions used in the authors' t-tests. What the authors describe as the "first appearance" of moralizing gods only occurs after writing or literate observers (Extended Data Figure 1), and is almost always preceded by a series of unknown values ('NA's), indicating no evidence that moralizing gods were present or absent (Table 1). Only one observation in the entire Seshat database--China's Middle Yellow River Valley--reports a known absence of moralizing gods in a world region before their "first appearance", and this lone data point is disputed by other historians[3].

	years until first appearance of moralizing god											
	-1000	-900	-800	-700	-600	-500	-400	-300	-200	-100	0	100
Deccan	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA
Kachi Plain	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1
Kansai		NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1
Konya Plain	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1
Latium	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1
Middle Yellow River Valley	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	1	1
Niger Inland Delta	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1
Orkhon Valley	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1
Paris Basin	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1
Sogdiana	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1
Susiana	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA
Upper Egypt	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	1

Table 1 | Moralizing gods across 12 key regions. Here, a "1" indicates moralizing gods are known to be present in the century-by-century data, "0" that they are known to be absent. An "NA" refers to "Unknown" moralizing god data in the authors' dataset. Generally, moralizing gods appear in Seshat simultaneously with or after the appearance of writing (green), contra ethnographic records of moralizing gods in non-literate societies (Supplemental S1.4, also see Extended Data Figure 1). Only one of the Seshat world regions includes a known absence of moralizing gods preceding a known presence (Middle Yellow River Valley, red, cf.[3]).

In total, 61% (n = 490) of all moralizing god observations passed to the statistical tests were originally labeled "Unknown" in the Seshat dataset (Extended Data Fig 2). Review of the authors' R scripts show that Whitehouse, *et al*. handled this problem with the moralizing gods variable by treating all cases of missing data ("NA") as known absences of moralizing gods ("0" or "FALSE") both in their *t*-tests (supplementary code file 04, line 39) and logistic regression (supplementary code file 06, line 48). The resulting correlation between missingness and moralizing god absence is r = 0.97, implying the study can be viewed as an analysis of the missingness patterns in the Seshat database (Supplemental S1.2).

These crucial decisions about the treatment of missing data are not reported in Whitehouse, *et al*.'s methodological description. In a public response, the authors have defended this re-coding on the grounds that, "the important distinction was whether or not there was direct evidence for the presence of such beliefs."[4] This is a restatement of a common fallacy, that the absence of evidence is evidence of absence. It is equivalent to assuming that the presence of moralizing gods in a society guarantees that society's religious beliefs will be recorded in historical datasets like Seshat (see Supplemental S1.3). As this is very difficult to justify, previous work in this literature using ethnographies limited such inferences only to cases where there are detailed accounts of a society's religious beliefs, but moralizing gods went unmentioned[5] (see Supplemental S1.4). In contrast, the four Seshat variables used to define moralizing gods, either as moralizing high gods (MHG) or any of three categories of behavioral supernatural punishment (BSP), are all labeled "Unknown" in the online Seshat databank for 490 of 801 historical observations used in the study (see Supplemental).

Further, because the definition of moralizing gods used by Whitehouse, *et al*. requires written evidence to be recorded, and missing data disproportionately come from small, non-literate populations below one million people, this assumption produces a powerful "forward bias" that pushes estimates of the adoption of such religious beliefs to a much later date in history. For example, Hawaii's population history is well-documented archaeologically, but the "first appearance" of moralizing gods reported in Whitehouse, *et al*.'s dataset is with the arrival of Europeans to write down those beliefs (see Supplemental data files). In light of Pacific ethnography, beliefs in supernatural agents who punish moral transgressions are likely to have appeared in Hawaii far earlier than post-contact accounts attest[3].

To test the sensitivity of Whitehouse, *et al*.'s "megasociety threshold" to their assumption that all missing values signify known absences of belief, we re-analyzed the Seshat data for the 12 key regions by moving the "first appearance" of moralizing gods back by one century–the smallest time unit in the data. Using the authors' analysis code, this minimal correction reverses their main result: moralising gods *precede* the dramatic rises in social complexity and the rate of increase in complexity is almost two times larger *after* the first documented appearance (Supplemental S2.1). While Table 1 illustrates that all 12 regions lack any data before the arrival of moralising gods, a softer version of this argument could claim that historically documented arrival of moralizing gods through conquest or mission is synonymous with their first appearance in the given region. However, even if we would assume that historical conquest and mission indeed documented the first appearance of moralizing gods (for which there are no data), re-running the authors' *t*-test for six geographical areas where moralizing gods did not arrive through conquest or mission shows compatible results as for the full 12 regions (Supplemental S2.1). These analyses indicate that Whitehouse, *et al*.'s key finding assumes the detection of the timing of moralizing gods's appearance in the Seshat was done with an extraordinary level of precision.

Though demonstrating the fragility of the "megasociety threshold", these tests still assume that all unknown values up to the first known data point in each region's time series represent the absence of moralizing beliefs. We thus revised the authors' logistic regressions using three standard approaches for missing data: analysis only on the complete cases, sampling from the known observations, and Laplace's principle of indifference[6-8]. In every result, the revised "takeoff" in the probability that moralizing gods are present occurs at much smaller population sizes than one million, before writing and major increases in social complexity (Fig. 1A, S6, Table S3, S4). These Bayesian models also permit prediction in eras without data, giving the probable emergence of moralizing gods between 600 and 1400 years before the "first appearance" recorded in the Seshat dataset (Fig. 1B, Fig. S4).

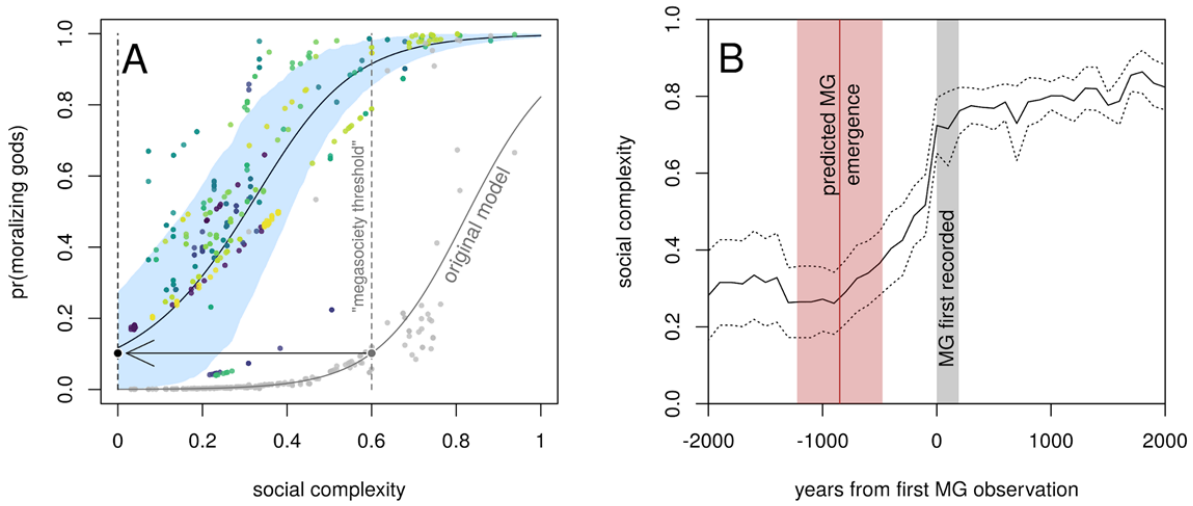


Figure 1 | Comparison of Whitehouse, *et al.*'s logistic regression model, where missing outcome data was coded as "moralizing gods absent", and a reanalysis removing unknown outcomes. Panel (A) Estimated relationship between the probability of moralizing gods being observed and social complexity, held at average distance and language similarity, fit on original dataset (NA's recoded as '0's; grey line) and reduced dataset that removed missing values (black line with blue 89% HPDI shading). Mean probabilities of "moralizing gods present" for the 490 historical observations with "unknown" outcome values are given as points: from the original model (grey) and grouped by NGA in revised model (each NGA is assigned a different color). Marked points indicate the "megasociety threshold" outcome probability in both models. **Panel (B)** Recreation of Whitehouse, *et al.*'s Fig. 2a, estimating forward bias only from known (non-NA) observations, now showing mean and 95% confidence interval for the predicted first emergence of moralizing gods, approximately 958 (SE: 210) years prior to their first observations in the Seshat database.

Among other statistical issues (see Supplemental and [7,9]) including concerns about data vetting and coding[3], these reanalyses cast serious doubt on Whitehouse, *et al.*'s main conclusion that moralizing gods appear only after rapid increases in social complexity globally. Given the problems of preservation of religious belief in the historical record, we conclude that the reported "megasociety threshold" is thus an artefact of Whitehouse, *et al.*'s decision to recode 61% of cases from missing data to known values, all indicating that moralizing gods "absent."

Author Contributions

Designed reanalysis: RS, BP, MM, ML, JH, RG, BB, QA

Performed re-analyses: RS, ML, BP, BB

Code review: RS, ML, BP

Review of historical coding and vetting: ES, WM

Wrote manuscript: all authors

Acknowledgments

We thank Whitehouse, *et al*. for responding to our request for clarification and to our early notification of our intent to submit this report. We applaud them for making their code, source material, and data public. We thank three anonymous reviewers for feedback, and also thank the following: the Department of Human Behavior, Ecology and Culture at the Max Planck Institute for Evolutionary Anthropology, Joshua Conrad Jackson, Richard McElreath, Elspeth Ready, Joseph Watts, Heidi Colleran, as well as Amir Ashtari, Adam Barnett and Tiffany Hwang for RA and administrative support.

Data & Code Availability

Supplementary data and re-analysis code are available at https://github.com/babeheim/moralizing-gods-reanalysis. Original analysis materials are available at https://github.com/pesavage/moralizing-gods. All software is freely available under Creative Commons License CC BY-NC-SA 4.0. Source materials available at http://seshatdatabank.info.

Declaration of Competing Interests

The authors declare the following competing interests: six co-authors (RG, JH, MM, WM, ES, RS) are involved in the Database of Religious History (DRH) project, another freely-available online historical database.

References

1.Whitehouse, H. *et al*. Complex societies precede moralizing gods throughout world history. *Nature* (2019). doi:10/gfw838

2.Turchin, P. *et al*. Seshat: The Global History Databank. Available at: http://seshatdatabank.info/ (2015).

3. Edward Slingerland et al., in press. "Historians Respond to Whitehouse et al. (2019), 'Complex Societies Precede Moralizing Gods Throughout World History,'"*Journal of Cognitive Historiography*.

4. Savage, P. E., et al. (2019). Reply to Beheim et al.: Reanalyses confirm robustness of original analyses [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/xjryt

5.Watts, J. *et al*. Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proc. Biol. Sci.* **282**, 1804 (2015).

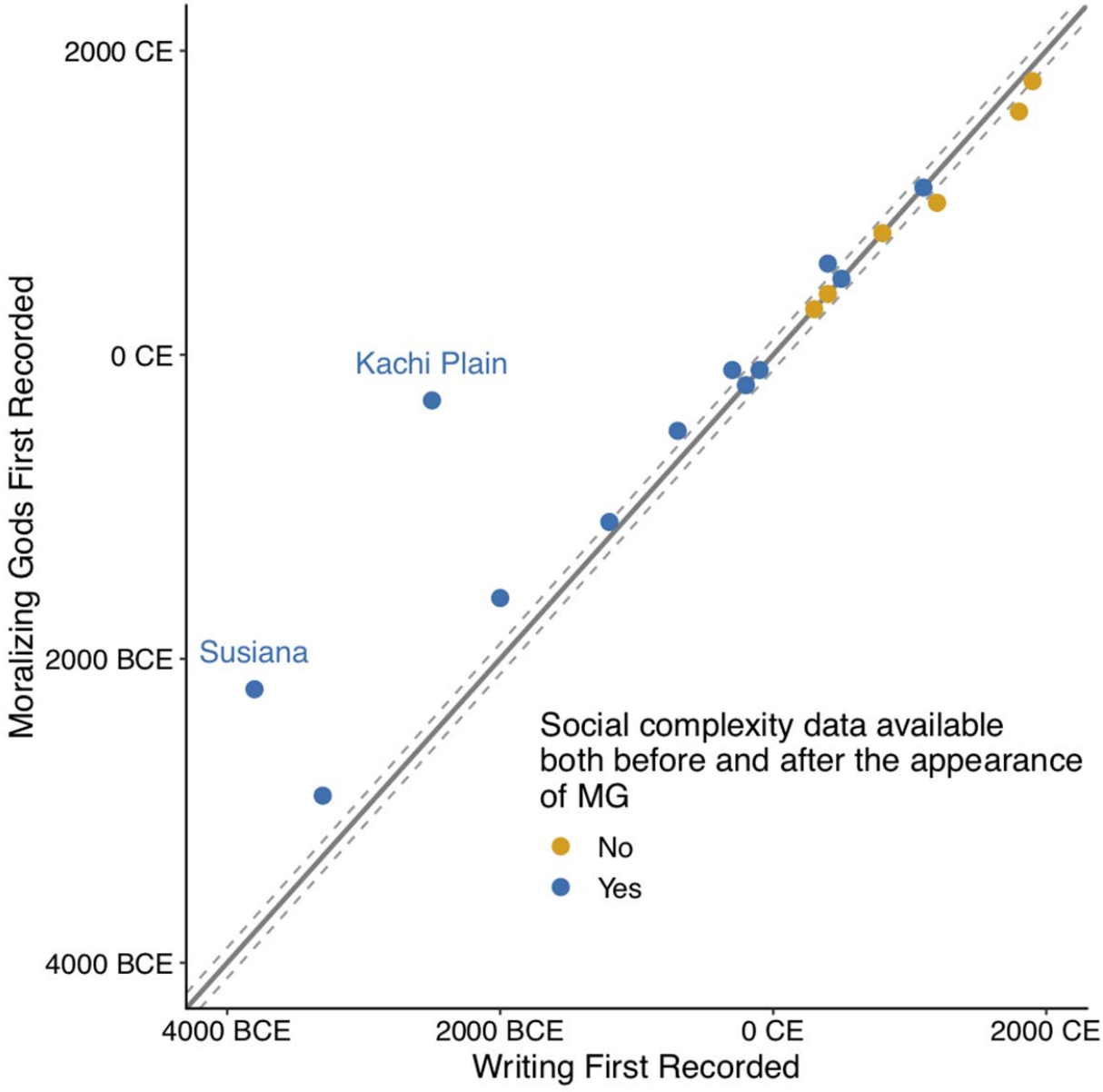
6.Little, R. J. A. & Rubin, D. B. *Statistical Analysis with Missing Data*. (John Wiley & Sons, 2014).

7. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan. (*Chapman and Hall/CRC, 2020).

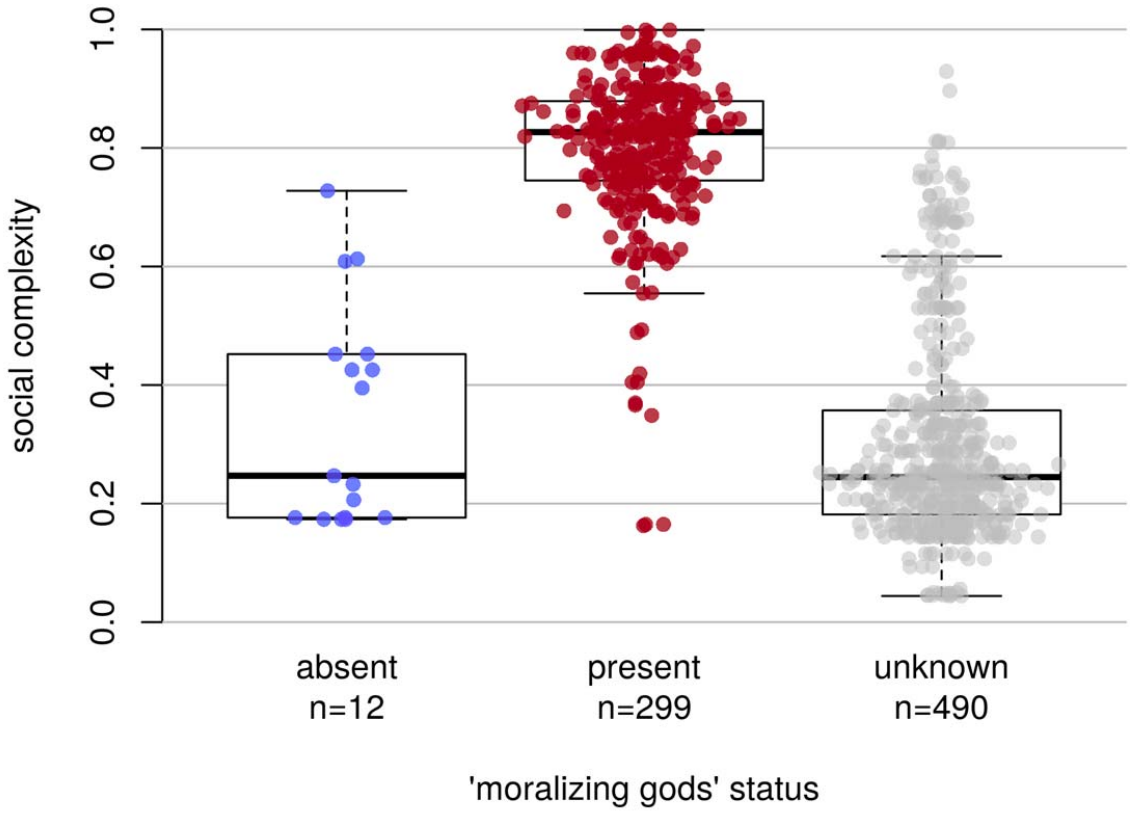
8. Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models*. (Cambridge University Press, 2007). doi:10.1017/CBO9781107415324.004

9. Shin, J., Price, M. H., Wolpert, D., Shimao, H., Tracey, B., & Kohler, T. A. (2019, May 17). From Scale to Information and Back: Human Social Evolution in the Holocene. https://doi.org/10.31235/osf.io/ejfbm

Extended data figures



Extended Data Figure 1 | The first appearance of writing and moralizing gods across NGAs. The solid line indicates when writing and moralizing gods are first recorded in the same century, and the dashed lines show when writing appeared 100 years before moralizing gods and when moralizing gods appeared 100 years before writing. NGAs are colored by whether social complexity data are available both before and after the appearance of moralizing gods or not. Only NGAs with social complexity data available both before and after the appearance of moralizing gods were included in the analysis (and only these NGAs are shown in Table 1). It must be noted that while writing first appears at 2500 BCE in the Kachi Plain, it is absent for the subsequent two polities in the dataset, and does not reappear until 300 BCE - the same time as the first appearance of moralizing gods.



Extended Data Figure 2 | Boxplots (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range) & distributions (data points) of "social complexity" score for N = 801 observations, by 'moralizing gods' outcome status. Before statistical analyses were performed in Whitehouse, *et al*., all "unknown" cases were treated as moralizing gods "absent" without explicit description in the manuscript.

Supplemental Materials

Treatment of missing data determines conclusions regarding moralizing gods

Bret Beheim, Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology
Quentin Atkinson, School of Psychology, University of Auckland
Joseph Bulbulia, School of Humanities, University of Auckland
Will Gervais, Department of Psychology, University of Kentucky
Russell D. Gray, Max Planck Institute for the Science of Human History, School of Psychology, University of Auckland
Joseph Henrich, Department of Human Evolutionary Biology, Harvard University
Martin Lang, LEVYNA: Laboratory for the Experimental Research of Religion, Masaryk University
M. Willis Monroe, Department of Asian Studies, University of British Columbia
Michael Muthukrishna, Department of Psychological and Behavioural Science, London School of Economics
Ara Norenzayan, Department of Psychology, University of British Columbia
Benjamin Grant Purzycki, Department of the Study of Religion, Aarhus University
Azim Shariff, Department of Psychology, University of British Columbia
Edward Slingerland, Department of Asian Studies, University of British Columbia
Rachel Spicer, Department of Psychological and Behavioural Science, London School of Economics
Aiyana K. Willard, Centre for Culture and Evolution, Brunel University London

Table of Contents

Introduction

In this supplement, we elaborate on the points made in the main text by critiquing and re-assessing three aspects of Whitehouse, *et al.*'s study[1]. In Section S1, we assess the authors' operational construct of "moralizing gods" in light of the field, data sources and ethnographic record. We also identify the major coding decision that drove the results. In Sections S2 and S3, we examine their statistical analyses, identify some issues, and provide more appropriate analyses. These analyses support opposite conclusions.

The authors' original code and history of subsequent edits can be found here: https://github.com/pesavage/moralizing-gods. Code and data for a full reproduction of the original analyses, all analyses contained in this response, and an html walkthrough of the analyses can be found here: https://github.com/babeheim/moralizing-gods-reanalysis.

S1. The problems of forward bias, missing data, and the ethnographic record

In this section, we discuss the problematic nature of Whitehouse, *et al*. assumptions concerning first appearance of crucial variables such as moralizing gods, as well as how these assumptions drive one of the more critical decisions in their analysis: the widespread conversion of missing data on moralizing gods to evidence of their absence.

S1.1 The problem of forward bias

The central outcome variable of their analysis is the first appearance of "Moralizing Gods" (MG) within the polities of a particular Natural Geographic Area (NGA). Here, MGs include both the traditional concept of "Moralizing High Gods" (MHGs) used in the Standard Cross-Cultural Survey and the Ethnographic Atlas, and the more inclusive category of forms of "Broad Supernatural Punishment" (BSP). An MHG is a creator deity who is "specifically supportive of human morality", such as in Judeo-Christian tradition. As many non-creator spirits are punitive and morally concerned, the authors have followed a more general trend towards inclusive concept of BSP in the literature, though they require that BSPs monitor behavior related specifically to fairness, reciprocity, and in-group loyalty. Effectively (and somewhat awkwardly), MG thus either refers to creator gods who care about morality in general, or non-creator spirits who specifically monitor these three domains of human cooperation. It is not clear why Whitehouse *et al.*'s coding didn't use the six other moral domains (harm, property rights, kinship obligations, respect, bravery, and purity).

Any analysis of the archeo-historical record that considers "first appearances," especially when comparing first appearance dates, must consider the problem of inherent forward biases. To understand why the first recorded dates of most cultural traits will be forward biased, consider what it takes for evidence of MGs to make it into the Seshat database:

1. A community must come to believe in a god or supernatural process that reliably monitors and punishes some moral transgressions related to (a) fairness, (b) in-group loyalty or (c) reciprocity . This is the true first appearance date.

2. Those transgressions must be related to fairness, in-group loyalty, or reciprocity but not other moral domains. Notably, Seshat also codes for six other moral domains, including murder and property crimes, but the authors ignored moralizing gods associated with these domains[1]. Here, first appearance dates will be pushed forward in time if the authors' three preferred domains happened to emerge after other domains, like murder.
3. These gods and their characteristics must have been accurately written down in sufficient detail for coding. This means that societies either have to first invent or otherwise acquire writing and then use it to express their beliefs about their gods punishing powers and moral concerns, or be described by missionaries, explorers or anthropologists motivated to accurately document these beliefs. The need for writing pushes first appearance dates forward until at least the invention/arrival of writing and the inclination to record supernatural beliefs (Extended Data Figure 1).
4. These possibly ancient written records have to survive to the present day. Such records need to be rediscovered, decoded and accurately described in secondary sources. The older an ancient record the less likely it is to both survive until the present day and be understood by modern scholars. This again biases first appearance dates forward in time.
5. Scholarly analyses of these writings have to be located by Seshat researchers and be accurately entered into Seshat. If early evidence is missed more frequently than later evidence, the first appearance dates are moved forward in time. See below for distinct cases in which Seshat researchers missed evidence of MGs earlier in history.

These and additional steps in the journey from a historical community's belief in an MG to the Seshat database can each contribute to pushing the first appearance dates in Seshat forward in time[2]. This influences first appearance dates in both absolute and relative terms, compared to the authors' measures of ritual and social complexity (which unlike beliefs, can often be inferred from non-written sources, such as monumental architecture, coins, etc.). This problem is further compounded by the authors' treatment of missing values in the historical record, described below.

[1] In other words, some punishing supernatural agents or forces may have been coded as 'absent' (MG = 0) because they were concerned about harming others (e.g., murder), being brave (e.g., in battle), violating property rights (e.g., stealing), and/or respecting local structures (e.g., "obedient to those above you in a hierarchy") but not the authors' three focal domains. This is even more worrisome in light of the historical record that suggests these coded-but-ignored features of moralizing gods are among the earliest documented supernatural concerns[2–6]. In Mesopotamia, for example, the earliest evidence of divine concerns we have identified comes from law codes that point to murder (harm) as the earliest concern in Ur-Nammu (2100 BCE)[7].

[2] Notably, the strength of forward bias will vary depending on the cultural trait in question. Both the Social Complexity and Doctrinal Rituals variables used by Whitehouse *et al*. would seem to suffer substantially less from forward bias compared to MG because they can be inferred from archaeological evidence. This means that any apparent differences in "first appearance" that emerge from historical data could merely represent differences in the strength of forward biases on different traits.

S1.2 Replacing missing values with known absences

Another central problem in the authors' argument is the treatment of unknown values ("NA") for MG in the analysis dataset. If Seshat's source files simply have no information regarding supernatural beliefs, entries appear as "unknown" (with expert review) or "suspected unknown" (coded "unknown", but lacking expert review). If either BSP or MHG are missing data for a particular time and place, MG will correspondingly be "NA" in the Whitehouse, *et al*. dataset, even if the other variable is absent ("0"). The requirements to code MG as absent ("0") are, in contrast, relatively strict: descriptions of a society's supernatural beliefs must exist, but both an MHG and all of the three required features for BSP must be absent as well ("0"). This outcome is correspondingly rare, as MG are known to be absent in only 12 observations, versus 490 observations categorized as unknown.

Whitehouse, *et al*. recoded these 490 missing data points from missing ("NA") to absent ("0"), in both their t-tests as well as their logistic regression. This conflates situations in which surviving records of supernatural beliefs exist but do not clearly describe MG, with situations in which no historical records exist, and we simply know nothing about the presence or absence of MGs either way.

In the t-tests, this occurs at line 39 of ESM code file 04, "prep_comparisons.r", which subsets the data to only those centuries for which MG are known to be present ("1"). However, by labelling the first such century the "first appearance" of MG, and calculating previous centuries as "pre-MG", the authors are treating missing data as equivalent to MG absence. Specific to the subset operation chosen, as described by the R manual on the subset argument, "missing values are taken as false" (R Documentation, v. 3.6.1). In their response to our critique, Savage, *et al*.[8] incorrectly state only the logistic regression performs this replacement.

In the authors' logistic regression, the substitution of missing data with known absences occurs at line 48 of ESM code file 06, "prep_regressions.r". This line explicitly replaces all missing values in every data variable with 0's, causing all subsequent appearances of the data to be indistinguishable from the "absent" code when passed to the logistic regression call.

S1.3 Implications of this assumption for the historical record

The implications of this replacement decision for small, pre-literate historical populations are important. The decision implies that no human population in the Americas of any size possessed moralizing deities before the arrival of European missionaries. Of the 390 observations in populations with a social complexity score less than 0.4, only 8 have any data concerning MGs, so the remaining 382 data points are assumed to be absent MGs.

In correspondence on this matter, Whitehouse, *et al*. have advanced the following assertion to justify this decision: "[G]iven the nature of the historical and archaeological record, if there was no evidence of moralizing gods we can treat them as being absent". In a more recent preprint[8], they appeal to the ethnographic record in a similar way: "Given this ethnographic evidence, we adopted a collection strategy that focused on identifying the earliest direct evidence for the presence of beliefs in moralizing gods, meaning that we treat such beliefs as absent until we have evidence of their presence. This strategy is reasonable because the first evidence of moralizing gods general post-dated the advent of written records". Because there are only 12 known absences and 490 missing values, mostly in small populations, this assumption is tantamount to assuming what the analysis seeks to test.

For any particular observation of the phenomenon of interest, let us distinguish two events. First, we define whether the phenomenon really is present in the study system for that observation, x, or absent, $!x$. Separately, we can also define the event that outcome data for that observation is recorded in our analysis dataset, d, or is unknown, $!d$. Our goal is then to estimate the true rate of occurrence of x across observations, the probability $p(x)$, given that $p(d) > 0$ and some information is incomplete.

Using this notation, we can write the probability that a phenomenon is present given that it is unknown in our data as $p(x|!d)$, which is distinct from $p(x)$, the probability the phenomenon was present, unconditional on whether the outcome is known or unknown in the dataset. The chance that the observation occurs, d, but is mis-coded from the true state, x or $!x$, is important but distinct from the missing data problem. We here assume that either the phenomenon recorded accurately or is unknown.

Expressed in the language of probability theory, the assumption that all missing values can be treated as information indicative of absence is

$$p(x|!d) < p(x) \tag{1}$$

Read literally, this says that conditioning on the outcome being unknown in our dataset, $p(x|!d)$, conveys information about the phenomenon itself which decreases the probability it is actually present in the system. By the law of total probability,

$$p(x) = p(x|d)\, p(d) + p(x|!d)\, p(!d)$$

It follows that,

$$p(x|d) > p(x)$$

Inverting the condition by Bayes' rule and simplifying, the result is

$$p(d|x) > p(d) \tag{2}$$

In other words, conditioning on the phenomenon's presence in the system signifies it is more likely to be recorded as data. With x representing the belief in MG in a given century and location, the assumption Whitehouse, *et al*. employ is an asymptotic version of condition (1),

$$p(x|!d) = p(x|!x) = 0 \tag{3}$$

where the absence of recorded data is equivalent information to the phenomenon's absence from the system. Applying the same logic to equation (3) as to condition (1) results in

$$p(d|x) = 1$$

By assuming (3), therefore, Whitehouse, *et al.*[1] also assume that if the phenomenon was actually present in a given century and location, it must have been recorded as such in their dataset. Or, equivalently, it cannot become missing data. This posits the existence of seemingly implausible mechanisms by which accounts of moralizing gods in, for example, a pre-literate South American pantheon would be recorded in the Seshat historical databank, while an equivalent cosmology without moralizing deities may be lost to time.

Correspondingly, the authors' logistic regression estimates (Extended Data Table 2 of the authors' original report) describe an implausible scarcity of MGs among small-scale societies. The reported logistic regression estimates imply that in observed societies with social complexity scores less than the median of 0.42 moralizing gods will appear in only about 2% of cases (SE: 1%, Fig. 1A) each century. Because social complexity closely tracks population size (Fig. S1), a practical interpretation is that the model predicts an *extremely low* chance of finding moralizing gods in any human populations with less than 50,000 people. As we discuss below, however, ethnographic descriptions from small-scale populations regularly contradict this prediction. In addition, such observations, if anything, provide only a lower bound for frequency of MGs in smaller scale societies.

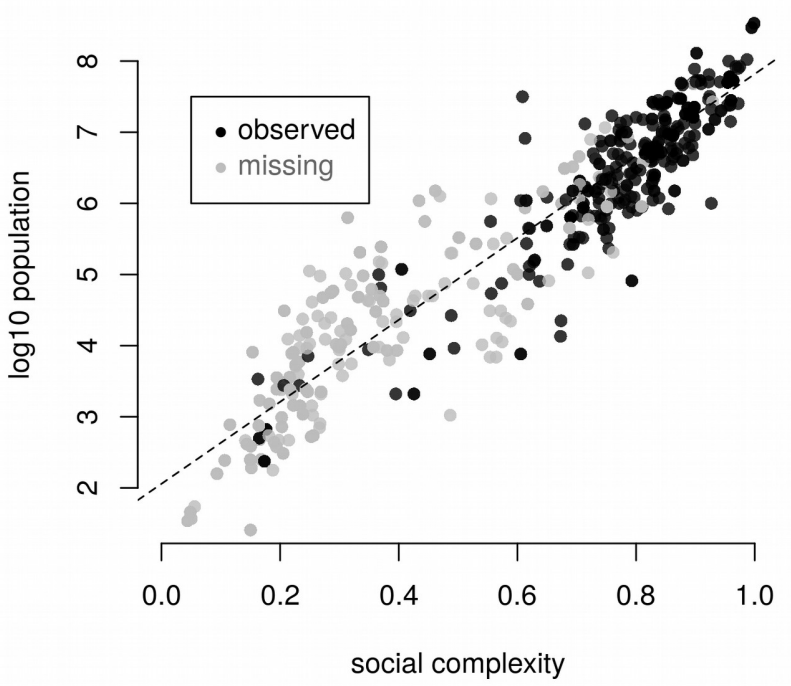


Figure S1 | Social complexity scores for *n* = 864 observations in the full Seshat dataset plotted against log-population sizes, with best-fit line (Pearson's $r = 0.94$) and shading indicating whether the MG outcome variable has known or missing values.

S1.4 Ethnographic evidence of moralizing gods

The fact that Whitehouse, *et al.*'s alleged first appearance dates for MGs depend heavily on the presence of writing (Extended Data Figure 1) presents a problem in light of the ethnographic record. The authors maintain that it is a "fact that evidence for moralizing gods is lacking in the majority of non-literate societies". Elsewhere[9], they claim that "Social scientists have long known that small-scale traditional societies–the kind missionaries used to dismiss as 'pagan'–envisaged a spirit world that cared little about the morality of human behaviour". Contrary to these claims (which inappropriately cite Bellah[10]), some of the most influential social scientists of the 20th century who worked on cross-cultural religious variation explicitly rejected such a strong position (see Evans-Pritchard 1965: 107-108[11]; Swanson 1974: 174[6]; Wallace 1966: 193[12]). Several quantitative anthropological analyses suggest a positive association between social complexity (i.e., "levels of jurisdictional hierarchy") and the presence of morally concerned high gods[6,12,13] (see below for further discussion of such data); as societies scale up, supernatural beings appear to become increasingly omniscient, punitive, and moralizing.

But these analyses also reveal non-trivial frequencies of supernatural punishment--including "moralistic" supernatural punishment--in small-scale and non-literate societies. In Boehm's review[14], for example, of 43 hunter-gatherer ethnographies covering 18 societies, there are instances of supernatural punishment of at least one behavior construed as "antisocial" and "predatory on fellow band members" in all groups. In Swanson's classic study[6] of 50 societies (78% of which had populations of 10,000 or more people and 78% of the sample had three or fewer "sovereign organizations"), only 6 (12%) report instances of having a "moralizing high god", yet 11 (22%) are counted as "uncertain". However, 11 have indicators of "active ancestral spirits" that "aid or punish living humans" and 28% (p. 14) "are invoked by the living to assist in earthly affairs". Moreover, all of the 50 sampled populations have some documented form of "supernatural sanctions for morality", that is, "behaviors that helped or harmed other people" (p. 212). In Watts, *et al.*'s Pulotu data set[15], 27 of the 74 (36%) Austronesian societies coded as "low political complexity" (acephalous or simple chiefdoms) had MGs[3]. If we expand Whitehouse, *et al.*'s targeted "moral" behaviors to include breaches of sexual mores, Brown's study[16] suggests that over a quarter of the sample (110 societies' from the HRAF) includes some reference to supernatural punishment. In fact, the first quantitative studies to systematically and directly ask large samples of individuals questions measuring the degree to which they believe traditional spirits know and care about various moral acts came about only very recently[17,18]. In these studies, responses were consistently well above zero. Notably, spirits' attributed concern and knowledge of moral behaviors was mediated by the proximity of those behaviors to the territories over which the spirits lorded (see below). The bottom line is that Whitehouse, *et al.*'s analysis—driven by their decision to recode all NA's as "0"—is inconsistent with the ethnographic evidence.

[3] Note that a project using this data reports that the authors coded their target variable (human sacrifice) "as absent (absent = 0) if ethnographic sources explicitly stated that human sacrifice was not practiced, or if there was no evidence of human sacrifice from a substantial description of the culture's religious practices."

Another problem with data used in analyses of human religions in non-literate societies is that much of it comes from early reports by missionaries and colonial administrators. Often, these individuals are evidently biased against seeing any elements that "look Christian" in pagan religions, thus leading to misleading information that later researchers subsequently code[4]. Additionally, qualitative data of this sort might be too ambiguous and therefore at risk of contributing to underrepresenting particular traits, especially if their absence in some cases is treated as a given in the coding process. Take, for instance, the case of the Orokaiva. Seshat[20] codes them as having no BSP or MHG (MG = 0) at any point in time. Whitehouse, *et al*. cite a source by Schwimmer[21], that contains the following depiction:

> If the Orokaiva, by and large, order their lives by the same moral principles, they would explain this by their common belief in certain demigods whom they all regard as their ancestors and as sources of authority, and who created certain institutions embodying moral norms to which they all subscribe. Not only do they obey the precepts of these demi-gods, they also re-enact their feats in ritual and identify with them during ceremonies, and in many of their regular expressive activities (p. 51).

These demi-gods "created certain institutions embodying moral norms to which they all subscribe." The Orokaiva themselves appeared to attribute their moral order to these gods, yet Seshat treats the presence of such gods as "absent". In Swanson's[10] dataset, the Orokaiva are coded as having no high god, but they do have ancestor spirits that "are invoked by the living to assist in earthly affairs" and have one coded instance of moralistic supernatural sanctions. Despite variation across sources, this example seems like a candidate for evidence of a MG, despite being treated as absent.

This does not imply that the gods of small-scale societies were no different from those of complex societies. Instead, the differences are often quantitative and related to the size of the sphere of supernatural punishment (from clan or regional members to all humans; see above), the particular domains of punishment (more of those supporting larger scale cooperation) and size of the supernatural incentives (e.g., contingent afterlives). Because of their own source biases, these quantitative anthropological studies only provide a lower bound on the percentage of societies with MGs and leave the question of the ubiquity of MGs unresolved[6,17-18,22–33]. Nevertheless, the ethnographic record provides no justification for the wholesale recoding of missing data as "MG absence" ("0") in less-complex, non-literate societies.

Given the strong evidence of MGs in non-literate societies and the authors' heavy reliance on written historical records for evidence of MGs, there is good reason to suspect a substantial forward bias, both in absolute terms and relative to measures of social complexity and doctrinal rituals. Below, we re-analyze the authors' data while taking seriously the challenge of forward bias and handling of missing data.

[4] Consider this missionary report[19] of the Abipón Indians of Paraguay, a source in the Standard Cross-cultural Sample (SCCS) and the Ethnographic Atlas (EA). The missionary describes his flock as follows, "the American savages are slow, dull, and stupid in the apprehension of things not present to their outward senses. Reasoning is a process troublesome and almost unknown to them. It is, therefore, no wonder that the contemplation of terrestrial or celestial objects should inspire them with *no idea of the creative Deity, nor indeed of any thing heavenly*" (p. 58; emphasis ours). These "slow, dull, and stupid" natives are nevertheless capable of conversion "when the good sense of the teacher compensates for the stupidity of his pupils" (p. 62). Based on this and two other sources, the SCCS and the EA code the Abipón as lacking a high god of any sort.

S2. Reanalysis of the pre/post-appearance comparison

In this section, we test the robustness of Whitehouse *et al.* findings and examine the implications of possible forward bias. We then demonstrate why Whitehouse *et al*.'s analytical approach was inappropriate, and finally we provide a more appropriate alternative analysis.

S2.1 Correcting for forward bias in the statistical approach of Whitehouse, *et al.*

When investigating the robustness of the causal analysis performed by Whitehouse *et al*., we found that the appearance of MGs in the archaeo-historical record usually occurs simultaneously with a sudden jump in social complexity. As illustrated in Fig. S2A, societies increased their complexity on average by 39% within the 100 years just before the appearance of MGs while the average between-century increase in social complexity for the preceding 700 years was approximately 7%. Whitehouse, *et al*. proposed that the correlation between an increase in social complexity and the appearance of MGs is an indication that a society has to pass a certain threshold of social complexity (> 0.6, which they call "megasociety") to evolve a religious system with moralizing gods. However, it is not clear why the appearance of MGs would require such a dramatic change in social complexity within a single century (Fig. S2A). The fact that MGs can be detected in the archaeo-historical record just after a sudden and unprecedented jump in social complexity is troubling and points to potentially hidden underlying factors that may have influenced both the measures of social complexity and researcher's ability to detect MGs presence.

We surmise that one of the explaining factors relates to the treatment of historical conquest by Whitehouse, *et al*.: at Deccan, Kachi Plain, and Sogdiana NGAs, the apparent rapid increases in social complexity and the appearance of MGs are a direct consequence of conquest by large empires (Fig. S2B). For instance, in 400 BCE, Deccan has a population of 20,000 with no MGs and no writing; yet, next century, the same NGA has a population of 18 million with MGs and writing, and then a century later reverts back to 20,000 people with no MGs and no writing. This dramatic change is caused by the fact that Whitehouse, *et al*. assign Deccan the population size, social complexity, and religious beliefs of another NGA, Kachi Plain during a century of Kachi Plain's imperial rule. Likewise, Kachi Plain acquires their MG when conquered by Susiana, suddenly increasing their social complexity from 0.33 to 0.90 within a single century. Sogdiana rises from a population of 10,000 in 500 BCE to 22 million in the next century when Susiana takes over their territory and imports MG.

Such a treatment of conquest results in the appearance of "sharp increases" in population size, social complexity and moralizing gods in the Deccan, Kachi Plain and Sogdiana NGAs. Given the analytical technique employed by Whitehouse, *et al*. (paired t-test), it appears as if these three NGAs spontaneously increased their social complexity within a century and only afterwards developed belief in MGs. In other words, the paired t-test treats the increase in social complexity for the conquered NGAs as coming before MGs (increase from year -100 to 0 in Fig. S2B) while, as a matter of fact, MGs arrived together with increase in social complexity due to conquest in year 0 on Fig. S2B. For example, Deccan has social complexity score of 0.31 one century before MG appearance (year -100); 0.88 in the century when MG appeared (year 0); and 0.31 one century after MG appeared (year 100). Using the equation for computing the rates of change in social complexity provided by Whitehouse *et al*. (see supplemental code file 04, lines 68-90), Deccan has a positive Pre-MG rate of social complexity change that is dramatically larger than the Post-MG change (0.57 vs -0.57). Moreover, the Post-MG change is negative and appearing as if the introduction of MG led to an extreme decrease of social complexity (while in fact it was caused by Kachi Plain leaving Deccan).

To test the robustness of Whitehouse, *et al*.'s results, we performed the same paired t-test analysis as they report in the original text (which we consider inappropriate, see below) but excluded the three NGAs with sharp increases in social complexity due to conquest. For the longest time-span of their data (2000 years Pre and Post-MGs), the results revealed that the five-fold higher rates of social complexity change between Pre- and Post-MGs reported by Whitehouse, *et al*. decreased only to a two-fold higher rate of social complexity change ($t = -5.28$, $df = 141$, $P < 0.001$). Treating the high rate of social complexity change related to conquest as if it would be there before the conquest (and hence, before the arrival of MGs) substantially inflates the reported results. Of course, the question of how to treat MGs brought by military expansion is difficult, especially when the historical records lack social complexity data specifically for the NGA after it had been conquered. A conservative approach that does not assume shifts in population size in millions within a century would be to exclude such military expansions from the analysis and focus solely on regions where MGs emerged independently.

Similarly, such a conservative approach might exclude also NGAs where MGs were brought by mission from other groups, as is the case for Kansai, Nigerian Inland Delta, and Orkhon Valley in the current data set. Figure S2C again shows mission is tightly correlated with a sharp increase in social complexity, although this increase is not as pronounced as in the case of conquest because each NGAs keeps its social complexity after mission rather than being assigned the social complexity of other groups. Nonetheless, Fig. S2C suggests that together with MGs, missionaries may bring writing (Nigerian Inland Delta, Orkhon Valley) and other cultural technologies that significantly enhance social complexity of the given NGA. Assigning the effects of these technologies on social complexity as if they would pre-date MGs in the paired t-test analysis biases the estimated causal relationship between social complexity and MGs, similarly as in the case of conquest. As illustrated in Fig. S2D, the six remaining NGAs where MGs were first recorded without mission or conquest from other groups display rather steady social complexity increase around the appearance of MGs.

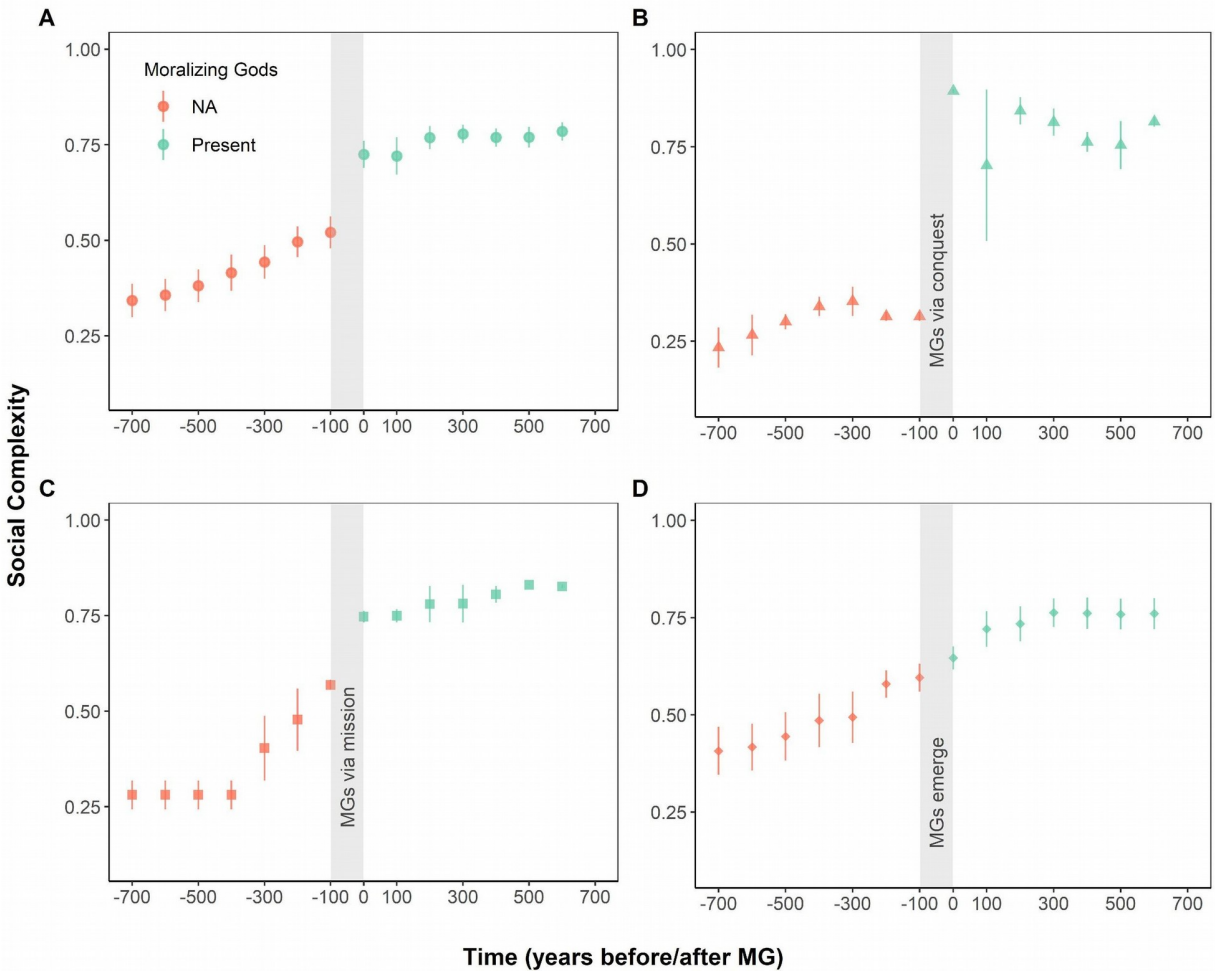


Figure S2 | Social complexity before and after the appearance of moralizing gods. Dots represent mean social complexity collapsed across NGAs. Bars represent +/- SE. The x-axis represents centered time before/after the presumed appearance of MGs at each NGA. Note that 0 on this axis represents widely disparate times, ranging from 2900 BCE to 1100 CE. **A.** The plot shows that MGs can be detected in the archaeo-historical records just after a sudden jump in social complexity that represents the smallest temporal unit in Whitehouse, *et al.*'s analysis (one century). **B.** The sudden jump in social complexity just before the appearance of MGs may be partially explained by the fact that 3 NGAs (Deccan, Kachi Plain, Sogdiana) were coded as having MGs only after these NGAs were integrated into larger empires with millions of inhabitants that already had MGs. **C.** Another proportion of the sudden jumps in social complexity may be explained by foreign mission (Kansai, Niger Inland Delta, Orkhon Valley). **D.** The remaining six NGAs where MGs were not first recorded through conquest by a larger empire or through mission show a steady rise in social complexity. Also note the sigmoid shape of the raw data typical for proportion data (see section S2.2).

Conducting the paired t-test analysis only for the six regions without conquest or mission revealed results congruent with the analysis of the sample comprising all 12 NGAs, but only for the longest time span of these six regions ($t = -4.50, df = 120, P < 0.001$); when we analyzed the 700 years timespan Pre- and Post-MGs where all NGAs have data (analogous to Whitehouse, *et al*. robustness checks), the results were much weaker ($t = -2.00, df = 48, P = 0.051$). This weaker result is partially caused by the decrease in degrees of freedom for the paired t-test analysis associated with a lower number of data points compared to the full span of Whitehouse, *et al*. data (2000 years Pre- and Post-MGs). However, it also shows that the results may not be as robust as claimed by Whitehouse, *et al*.

Importantly, even for the six regions where MGs were first recorded through conquest or mission, there are no data on MGs presence/absence pre-conquest/mission (see Table 1 in the main text). This suggests that first MG detection crucially depends on the presence of writing, be it writing of the conquering/missionizing NGA or writing of the NGA where MGs emerged without conquest/mission. As a consequence, the dependence of MG detection on written records naturally biases the first MG appearance toward more recent dates, a phenomenon we call forward bias. This problem is illustrated in the main text by the example of Hawaii, where Seshat treats the first appearance date of MG as the date when the first European Explorers arrived and documented Hawaiian beliefs. Another illustration is the Mesopotamian sun god Shamash, coded by the Seshat team as an "active" god when he first appears in writing (2250 BCE). Shamash, however, appears in iconography at least half a millennium (2750 BCE) before he appears in writing[34]. Not coding Shamash as "active" in 2750 BCE requires assuming that people did not think of Shamash as an active god, participating in their lives, until he happened to enter the preserved written record—a rather unlikely assumption, especially considering the fact that some sort of organized and widespread belief in Shamash and his power must have motivated the creation of iconography in the first place. Furthermore, even without such pre-existing material evidence, it is reasonable to assume a time-lag between the spread of religious beliefs and their first appearance in writing.

Note that Whitehouse, *et al*. tested their results for robustness against dating uncertainty by randomly placing MG appearance within the time-span of the polity in which MG first appeared.[5] However, for 11 out of 12 NGAs, MGs were always detected in Seshat during the first century of the polity's existence, so any random placing of MGs within the polity time-span would always make the first appearance of MG more recent, i.e. only worsening the forward bias. In other words, in Whitehouse, *et al*.'s robustness check, there was almost no possibility that MGs might have appeared earlier, only a possibility that they appeared more recently.

[5] While the analysis code for this robustness check was not part of the code made available by the authors, we infer on this robustness check from the Methods section in the original manuscript: "Our primary analysis treated moralizing gods as being present from the beginning of the polity in which they appeared. To ensure that our analyses were not affected by dating uncertainty, we reran the analyses randomly resampling to treat moralizing gods as appearing at some point from within the full date range of this polity (for example, 2900–2700 BCE for Egypt)."

To examine the effects of possible forward bias, we moved the first appearance of MG at each of the 12 NGAs 100 years back, the smallest possible correction given the resolution of the original data. Using the same analytical techniques as in the original paper, the results of the paired t-test for the longest time span of the data (+/- 2000 years) now shows that MGs positively predict the rise in the rate of social complexity ($t = 4.04$, $df = 201$, $P < 0.001$) -- not a drop in the rate of social complexity as Whitehouse *et al*. find. Furthermore, moving the first MGs appearance 300 years back (still a very conservative estimate, see Fig. 1B), the rise in the rate of social complexity change after the appearance of MGs is even stronger ($t = 6.18$, $df = 203$, $P < 0.001$). Qualitatively similar results can be reached by using the shorter time span of the data (+/- 700 years; see Table S1), indicating that the conclusion about the role of MGs in social complexity increase reached by Whitehouse *et al*. is extremely sensitive to small changes.

Nevertheless, the results of this forward bias analysis are driven by the problematic treatment of conquest and mission described above, only in the opposite direction. That is, moving MG appearance 100 years back at each site treats social complexity increase due to conquest and mission as if it was caused by MGs. We therefore performed the same forward bias analysis, but limited the number of NGAs only to the six NGAs where moralizing gods were not recorded through conquest or mission. The results of the forward bias analysis for these 6 NGAs again document the fragility of Whitehouse *et al.*'s main conclusion: moving MG appearance 100 years back for the longest time span of the data shows much weaker results (*t = -2.61, df = 111, P = 0.010*) and these results are no longer significant at conventional levels (*P < 0.05*) for the 700-year timespan of the data (*t = -0.46, df = 41, P = 0.645*). Moreover, moving MGs appearance 300 years back again reverses the results reported by Whitehouse *et al*.: MGs positively predict the rise in the rate of social complexity both for the longest time span of the data (*t = 4.20, df = 121, P < 0.001*) and also for the 700-year timespan (*t = 2.74, df = 41, P = 0.009*). For an overview of the t-test results see Table S1. This reanalysis demonstrates that, for their causal proposition to hold, researchers would have to be able to detect the first appearance of MG beliefs in the archaeo-historical records with a precision of +/- 50 years and assume that people started to write about religious beliefs immediately after their appearance. We regard this as unlikely.

	Original result (n = 12)		Conquered NGAs excluded (n = 9)		Conquered and missionized NGAs excluded (n = 9)		Forward bias correction by 100 years (n = 12)		Forward bias correction by 300 years (n = 12)		Forward bias correction by 100 years (n = 6)		Forward bias correction by 300 years (n = 6)	
	t-test	SC rate	t-test	SC rate	t-test	SC rate	t-test	SC rate	t-test	SC rate	t-test	SC rate	t-test	SC rate
+/-700	-4.64***	5.33	-3.21**	1.99	-2.00	1.64	3.61***	0.39	4.02***	0.45	-0.46	1.10	2.74**	0.45
+/-2000	-6.44***	5.03	-5.28***	2.11	-4.50***	2.00	4.04***	0.50	6.18***	0.45	-2.61*	1.39	4.20***	0.50

Table S1 | An overview of paired t-test results assessing the robustness of Whitehouse *et al*. results.
social complexity rate is the rate of change in social complexity Pre-MG divided by the rate of change in social complexity Post-MG.*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

S2.2 Growth curve models of social complexity

Putting the problems of forward bias and flawed treatment of missing values aside, we also found their analytical approach inadequate to assess the complex causality between MGs and social complexity. Treating MGs as an exogenous intervention that begins to influence social complexity at some year "0" and did not exist before disregards the obvious likelihood that if there is a relationship, MGs would have co-evolved with social complexity through a complex feedback-loop process rather than suddenly appearing. However, even if we do assume that MGs suddenly appeared at a specific point in time without prior interaction with social complexity, the paired t-test employed by Whitehouse, *et al*. remains inadequate for the nature of the analyzed data. The computed rates of social complexity change analyzed with the t-test include 400 data points (i.e., 200 Pre-MG time-points and 200 Post-MG time-points), but ignore the fact that these data points are nested within 12 focal NGAs. Additionally, some NGAs have more observations than others (ranging from 0 to 13 missing centuries per NGA). This data structure warrants a model that is flexible enough to handle repeated measures through space (polities within NGAs) and time. When comparing the rates of change for specific time windows (e.g., 100 years Pre/-Post-MG), however, the paired t-test analysis treats each pair of data-points as an independent observation. In other words, it only considers one time-window (e.g., 100 years Pre/-Post-MG in Susiana) to be a repeated measure, while the other time-windows (e.g., 200 years Pre/-Post-MG in Susiana) are considered to be from a different "individual" (while actually being from the same NGA and likely very similar to the 100-year time-window). This approach severely violates the assumption of independence[35] and artificially inflates the degrees of freedom for the t-test. In our view, the data have at least two hierarchical levels corresponding to their nesting within NGAs and their further nesting within world regions. We built a linear mixed model accounting for this nesting structure; however, the goodness-of-fit assessment of this model revealed severe deviation from the normality assumption (see Supplementary R code for diagnostic checks and plots).

To resolve the violation of the independence and normality assumptions, we used a multilevel growth-curve model that accounts for data-nesting and affords flexibility in modeling the distributional assumptions. While this model is still too crude for modeling complex causal relationships (e.g., it cannot model continuous feedback between the growth of social complexity and MGs), it produces precise estimates based on the assumed data-generation process[36,37] rather than a simple test of difference. To fit the growth curve model of differences in social complexity Pre- and Post-MG, we used raw social complexity as the outcome variable (rather than already pre-calculated rates of social complexity change); time, MG presence, and their interaction as predictor variables; NGA and world region as nesting factors for fitting varying intercepts; and a varying effect of time for each NGA to account for the NGA-specific rate of social complexity change[6]. Finally, to account for the fact that the social complexity data were scaled between 0 and 1, we used a beta distribution that estimates the mean and dispersion of scaled social complexity data, which are typically heteroscedastic and skewed[38].

---

[6] The full model is defined below, where g is the logit link for beta regression; Y_{ijk} is social complexity at time-point *i* within NGA *j* and world region *k*. β_0 is a fixed intercept, u_{0j} is a varying intercept for NGA *j*, and u_{0k} is a varying intercept for world region *k*. β_1 is the parameter for the fixed effect of time, u_{1j} is the parameter for varying effects of time across NGA *j*, and X_{1ijk} is the value of the time-point *i* for NGA *j* and world region *k*. Analogically, β_2 is the parameter for the effect of Pre-/Post-MG and X_{2ijk} the value of Pre-/Post-MG. β_3 is the parameter for the interaction of Time*Pre-/Post-MG, and ε represents the error term for the assumed beta distribution with parameters μ representing location and ϕ representing dispersion:

$$g(Y_{ijk})=((\beta_0+u_{0j}+u_{0k})+(\beta_1+u_{1j})X_{1ijk}+\beta_2X_{2ijk}+\beta_3X_{1ijk}X_{2ijk}+\varepsilon_{ijk})\sim \text{Beta}(\mu,\phi)$$

This model allows us to examine the change in social complexity before and after the assumed MG arrival while adjusting the model for the various nestings presented in the data as well as for the assumed beta distribution. Indeed, goodness-of-fit measures indicated that this full model fits the data reasonably well (see Supplementary R code). Table S2, Model 1 displays the results of the full model for the period of +/- 2000 years while Model 2 displays the results of the full model for the period of +/- 700 years. In these models, Time is the estimated increase in social complexity over one century before the arrival of MGs; MG [Pre vs. Post] is the difference in intercepts for the Pre-MG and Post-MG regression lines, i.e., social complexity 700/2000 years Pre-MG vs. social complexity at the time of the supposed MG appearance; and Time by MG interaction is the difference in slopes for the Pre- and Post-MG periods[7].

The results of the multilevel growth curve model reveal that by using a more appropriate statistical approach, the original t-test results presented by Whitehouse, *et al*. do not hold (see Tab. S2). While the time variable in both models indicates that the Pre-MG rate of social complexity growth was positive (more so in Model 2), the slopes of the Post-MG rate of social complexity growth were not significantly different at conventional significance levels ($P < 0.05$; albeit, the 95% CIs in Model 2 only slightly overlapped 0). Transforming the beta estimates from the logit link revealed that in Model 1, the average estimated rate of social complexity change per century Pre-MG was 0.01 and Post-MG was 0.006; and in Model 2, the average estimated rate of social complexity change Pre-MG was 0.03 and Post-MG was 0.007.[8] See Fig. S3 for plotted estimates over the raw data. While the slope difference in Model 1 is negligible, the model for the 700-year timespan estimates large differences in the rate of social complexity change Pre- and Post-MG. However, these estimates are quite uncertain given the low number of NGAs ($n = 12$) that served as a hierarchical factor in the model and, therefore, likely inflated. Furthermore, the problematic treatment of conquest discussed in section S2.1 caused the Post-MG growth rates to start very high on a bounded scale (intercept for the Post-MG social complexity was 0.74 out of 1) and given the nature of proportion data (sigmoid shape, see Fig. S2D), there was much smaller room for growth compared to the Pre-MG social complexity starting at 0.33 (see Fig. S3B).

[7] Note that our regression approach does not analyze the 100-year period between the last time-point without MG and the first time-point with MG. While the *t*-test analysis Whitehouse, *et al.* performed treats these 100 years as 'Pre-MG', such a step has to assume that MG appeared suddenly and *ex nihilo*. We allow MGs to arise within the 100 years, still an unreasonable assumption but necessary to remain consistent with the original analyses.

[8] However, note that the change in growth rates is not linear (it was back-transformed from the logit link), hence the averages are slightly misleading. For full growth rates see Fig. S3.

	Model 1 (2000-year timespan)	Model 2 (700-year timespan)
Intercept	-1.08***	-0.70***
	(-1.38, -0.78)	(-0.94, 0.46)
Time	0.04***	0.13***
	(0.02, 0.07)	(0.07, 0.18)
MG [Pre vs. Post]	2.15***	1.74***
	(1.93, 2.37)	(1.43, 2.094)
Time*MG	-0.008	-0.08
	(-0.03, 0.01)	(-0.16, 0.006)
Observations	429	168
NGAs	12	12
World Regions	6	6

Table S2 | Estimates with 95% CI from the models of social complexity. Estimates are raw beta estimates with the logit link. Time is social complexity change per millennium. Transformed estimates are plotted in Fig. S3. *P < 0.05; ***P < 0.001.

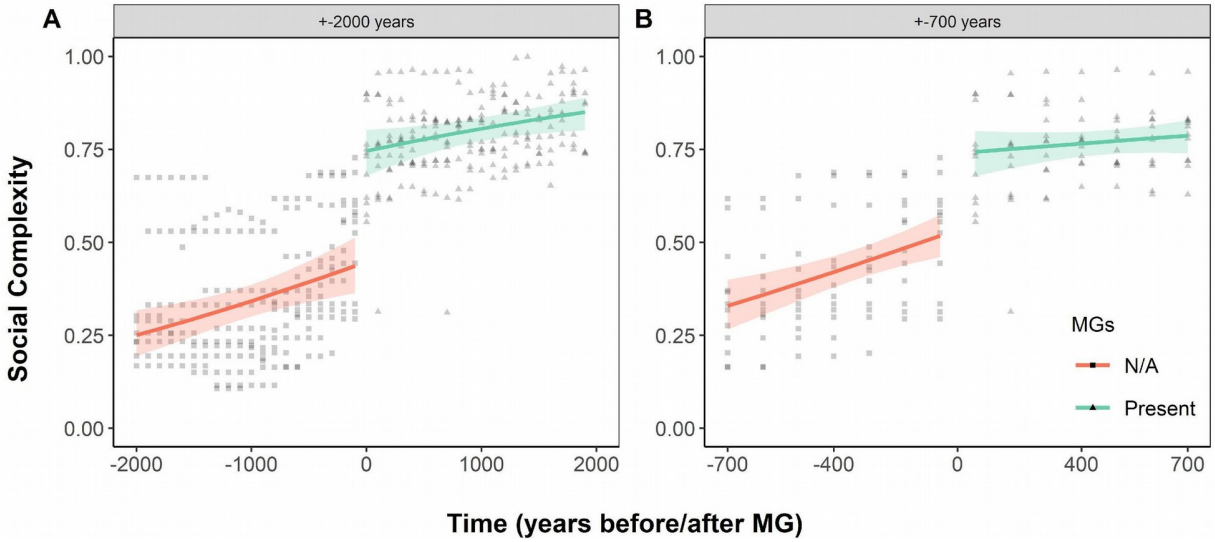


Figure S3 | Estimated regression lines with 95% CI from models in Tab. S2 plotted over scatterplots of raw data. The two plots differ by the timespan of analyzed data (either 2000 or 700 years Pre- and Post-MGs). Note that 0 on this axis represents widely disparate times, ranging from 2900 BCE to 1100 CE. Also note the sudden jump between Pre- and Post-MG regression lines produced by the way Whitehouse *et al.* treat conquered NGAs (section S2.1).

In summary, we believe that given the structure of the data used by Whitehouse, *et al.*, the growth-curve model represents a more appropriate and nuanced analytical approach. If the original data would not assume that the absence of evidence is evidence of absence (see below), this model could provide more reliable estimates compared to the paired t-test while respecting the test's assumptions. The

growth-curve model also allows for further investigation of non-linearity in growth curves (and their Pre-/Post-MG difference), which we omitted from the current analysis for the sake of simplicity[9].

S3. Reanalysis of the logistic regression model

Here we examine Whitehouse, *et al*.'s use of logistic regression to predict the appearance of moralizing gods (their Extended Data Table 2). First, we reproduced Whitehouse's essential results using the materials made public by the authors, including their figures, regression estimates, and data sample sizes. We then examined the sensitivity of their key results to the assumption that missing values (NA's) in their outcome variable could be converted to the absence of moralizing gods' (rewritten as "0"s). Because missingness in the Whitehouse, *et al*. data strongly correlates with smaller population sizes and lower social complexity (Extended Data Figure 2; Section S1.2), their conversion of all missing values to 0's is an extremely favorable assumption for the authors' preferred conclusion. Indeed, we find that one of their main results -- that increases in social complexity precede moralizing gods -- entirely hinges on this assumption, and for each reasonable alternative we see the opposite pattern.

We here consider the pattern of missingness as a "missing at random" scenario[47], for which the most principled approach is to simply drop the missing values. Despite the authors concerns that outcome missingness is clearly a function of a population's social complexity[48], estimates of the regression intercept and effect of social complexity will be unbiased so long as the model conditions on social complexity as a predictor variable.

The resulting analysis excludes 490 observations with missing values on the outcome variable. Because there are two lag terms in the published model, which use the binary outcome values from the two previous time periods within each NGA, two additional complications appear. First, for the earliest observations where the outcome value is known, we will have missing values in the lagged terms we must deal with. Second, because no world region shows any within-region variance in the outcome value without the "NA to 0" imputation (except one observation in Middle Yellow River Valley, cf. [49]), the lag terms become perfectly correlated with the outcome, which no longer allows for a logistic regression time-series approach. To maintain comparability, we removed the lag terms and re-estimated the probability of moralizing gods' appearance as a logistic regression on social complexity, distance, and language similarity as calculated in the original analysis. To account for non-independence across observations by region, we added varying-effect intercept terms for each region, with mildly regularizing priors on each parameter. Specifically, each logistic coefficient is given a Gaussian prior with mean 0 and $\sigma = 4$ log-odds units, with a Gaussian centered intercept prior of 0 with σ of 1 log-odds unit fit by Hamiltonian MCMC with the rstan package[50].

[9] The analyses presented in this section were conducted using R, version 5.3.5[39] and packages DHARMa[40], dplyr[41], glmmADMB[42], glmmTMB[43], lme4[44], reshape[45]. The figures were plotted with the help of ggplot2[46] package.

These changes alter the estimates of covariates like spatial proximity, but when fit on the full dataset preserve the essential pattern of rapid increase in the probability of moralizing gods around a social complexity score of 0.6 (Table S3). When judged solely on the MG data that is not missing, however, this revised model finds that the relationship between social complexity and moralizing gods is far weaker. The estimated frequencies of moralizing gods predicted by this model, particularly in smaller-scale societies with lower levels of social complexity, are much more consistent with the ethnographic record. For example, the contrast in Figure 1A is striking for low and intermediate levels of social complexity: societies with an social complexity of 0.4 are predicted to have essentially no moralizing gods when NA's are recoded as zeroes (per Whitehouse, *et al*.) but over half are expected to have moralizing gods when NA's are removed.

	Reduced Dataset *All 'NA' removed*		Full Dataset *All 'NA' to 0*	
	Est. (SE)	P(sign)	Est. (SE)	P(sign)
Intercept	1.05 (0.65)	0.05	-1.48 (0.58)	0.01
Social Complexity	8.64 (2.33)	<0.01	15.11 (1.56)	<0.01
Phylogeny	1.40 (3.88)	0.37	5.08 (2.98)	0.04
Space	6.82 (2.55)	<0.01	3.22 (1.32)	0.01
NGA Varying Effect	2.31 (1.03)	-	2.57 (0.68)	-
N	336		801	
Deviance	104.7		281.2	

Table S3 | Revised logistic regression estimates for the presence or absence of moralizing gods in the reduced dataset and full dataset, with means, standard errors, and probability each effect is null or negative (the Type-S sign error)[44]. Outcome values coded as "unknown" or "suspected unknown" in the Seshat database and NA in the analysis dataset were removed, and to account for within-region non-independence, a varying-effects intercept was added for each NGA. Without NA values converted to 0, lag terms in the original model become linearly dependent with the outcome variable and are removed as well. Social complexity is centered on 0.5 to aid intercept interpretability.

With this revised model, we can then infer the "first emergence" of MGs for each NGA here defined probabilistically as the earliest point at which 80% of the posterior probability mass is above $P = 0.5$ on the outcome probability scale, which indicates reasonable certainty that moralizing gods are present conditional on the available information (Figure S4). Combining these estimates as in the original analysis provides an average emergence point of approx. 1000 years before first observation in Fig. 2B. For comparison, Fig. S5 shows similar posterior probability calculations drawn from Whitehouse, *et al*.'s model fit on the original dataset including all NA's as known 0's. Consistent with the corresponding counterfactual predictions in Figure 1A, this model estimates the probability of MG emergence to be close to 0 for every focal NGA until moralizing gods are actually observed.

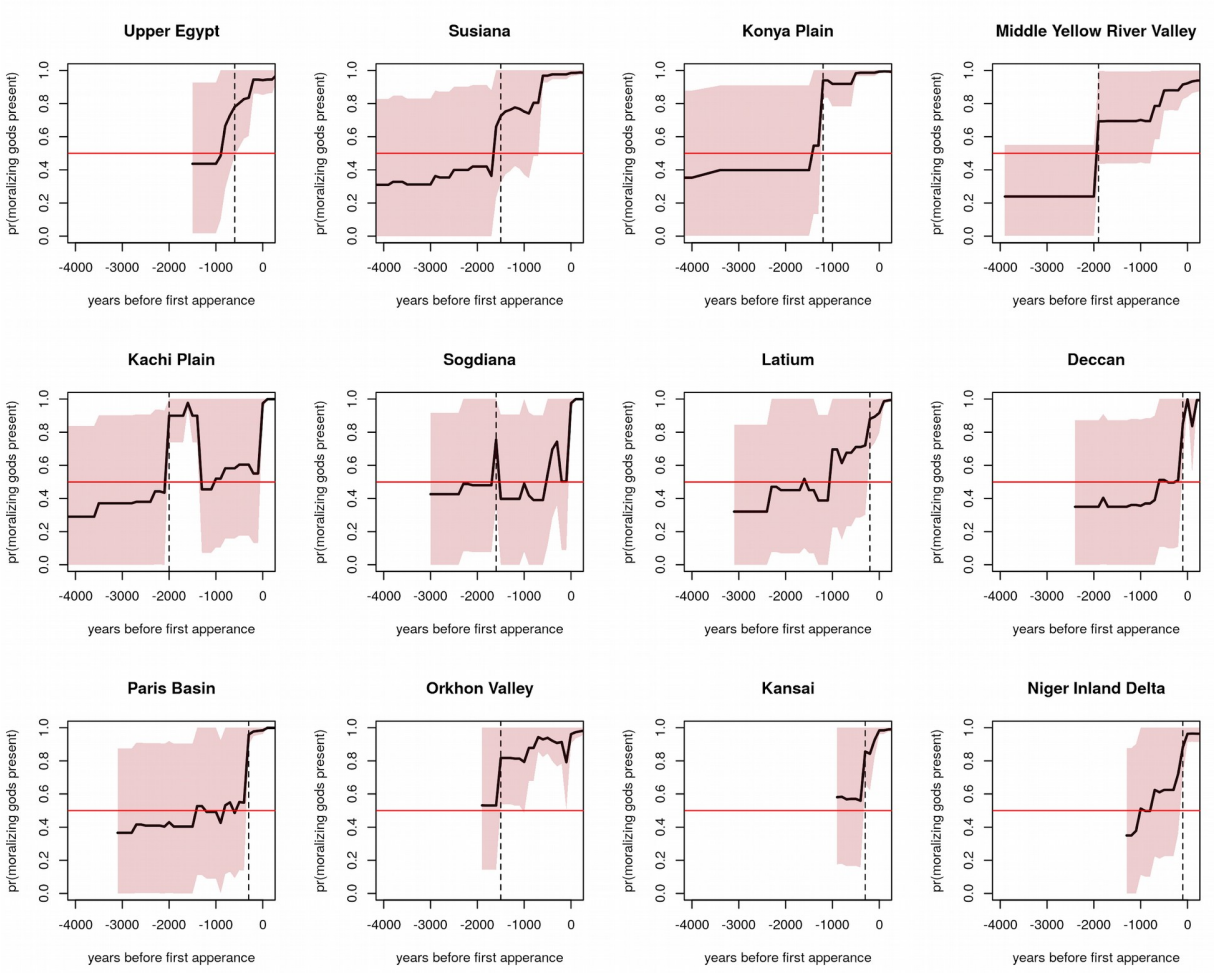


Figure S4 | Posterior predictions for the probability of moralizing gods present by year for Whitehouse, *et al*.'s main 12 NGAs in their analysis, drawn from the model described in Fig. 1A and Table S3 measured in years before their first documented appearance in the Seshat database. Posterior mean probability (black line) accompanied by 89% HPDI (red shading) indicates a high chance of MG presence in every site several centuries before recorded first appearance. Dashed lines indicate the first year at which 80% of posterior mass is above a probability of 0.5 (coin flip), used as a rough estimate of the "first emergence" of moralizing gods in Fig. 1B.

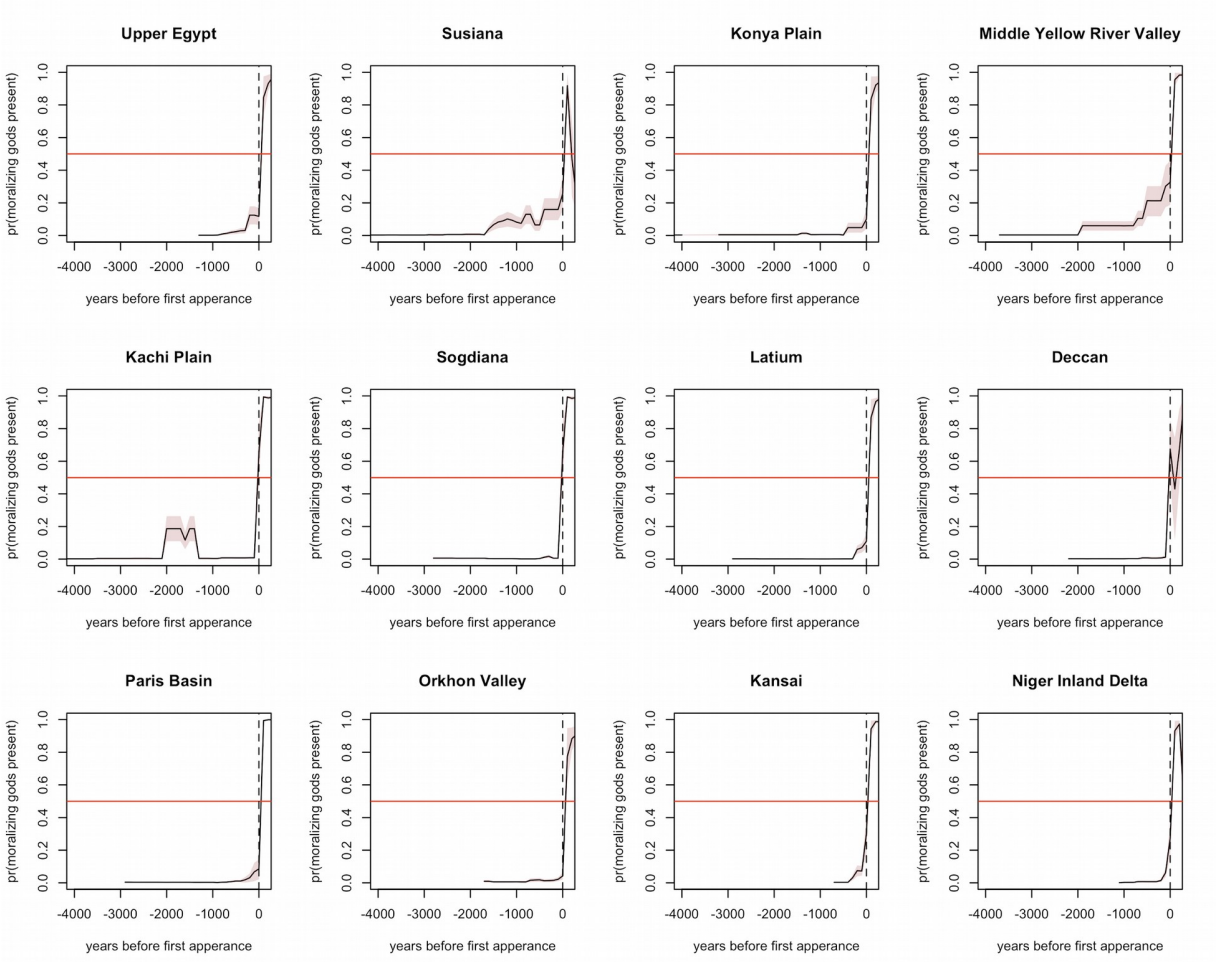


Figure S5 | Posterior predictions for the probability of moralizing gods present by year for Whitehouse, *et al*.'s main 12 NGAs in their analysis, drawn from their original regression model, measured in years before their first documented appearance in the Seshat database (dashed lines). Posterior mean probability (black line) accompanied by 89% HPDI (red shading) predicts almost no chance of moralizing gods, in contrast to Figure S4.

While reducing to complete cases (i.e., removing outcome "NA's") is a standard solution in this situation, it is important to consider alternative imputation methods that are more conservative against the hypothesis favored by Whitehouse, *et al*. For example, they may have instead assumed that moralizing gods appear just as often in small, non-complex populations as in large, complex ones, implying an imputation rule of randomly assigning 1's to missing values at the same rate of occurrence (96%) as in observed values[31]. Or, citing Laplace's "Principle of Indifference", they might have considered how, absent any knowledge of the features of a non-literate society's cosmology, we are equally ignorant of their presence or absence, and assign 50% of missing values "1", and 50% "0". The resulting regression coefficients by these missingness models are presented in Table S4. In both alternative cases, shown in Table S4 and Figure S6, we do not see the rapid increase in the probability of MG appearing after societies have become large and complex megasocieties.

	Original Model *All 'NA' to 0*		Alternative 1 *96% 'NA' to '1'*		Alternative 2 *50% 'NA' to '1'/'0'*	
	Est. (SE)	P(sign)	Est. (SE)	P(sign)	Est. (SE)	P(sign)
Intercept	-3.09 (0.36)	<0.01	1.30 (0.48)	<0.01	0.12 (0.20)	0.27
Social Complexity	9.78 (1.42)	<0.01	1.08 (0.89)	0.11	2.81 (0.46)	<0.01
Lag1	3.86 (0.67)	<0.01	1.51 (0.47)	<0.01	0.25 (0.18)	0.09
Lag2	0.83 (0.66)	0.11	0.56 (0.50)	0.13	0.85 (0.19)	<0.01
Phylogeny	7.92 (3.97)	0.02	8.25 (5.15)	0.05	6.60 (3.44)	0.02
Space	-2.03 (1.15)	0.04	0.26 (1.23)	0.42	0.85 (0.66)	0.10
N	801		801		801	
Deviance	169.0		267.8		778.4	

Table S4 | Regression estimates for the presence/absence of moralizing gods under three "missingness regimes" for the 490 missing values: means, standard errors, and posterior probability each effect is null or negative (the Type-S sign error)[51]. The "Original model" treats all missing values ("NA") as 0, corresponding to the estimates in Whitehouse, *et al*. Two alternatives using the same regression model, but different imputation methods for missing values: (1) 96% of NA's assigned randomly to '1' (the frequency of occurrence in the observed data), and (2) 50% of NA's randomly assigned to '1', 50% to '0'. Social complexity is centered on 0.5 to aid intercept interpretability.

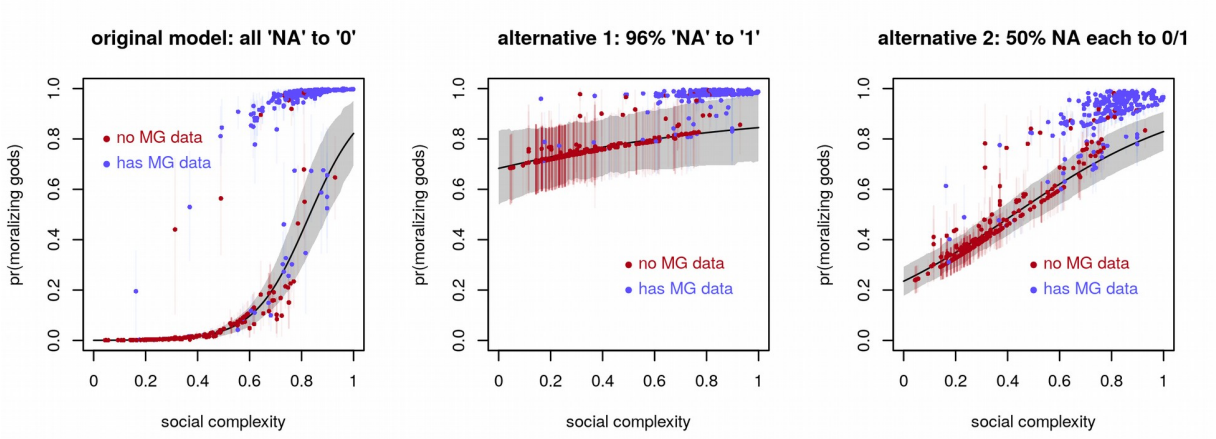


Figure S6 | Estimated relationship between the probability of moralizing gods being observed and social complexity, for original model and two alternatives described in Table S4. Trend lines (black) are all held at average distance and language similarity, with 89% HPDI shading. Mean probabilities of "moralizing gods present", for both "known" and "unknown" outcome values, are given as colored points with 89% HPDI intervals.

Though we prefer the missing-at-random (MAR) approach (removing the cases with NA's) to either of our Alternatives 1 or 2, these other approaches to dealing with missing data nevertheless illustrate how difficult it is to arrive at Whitehouse, *et al*.'s preferred result--moralizing gods post-date the rise of complex societies--using Seshat data. To summarize, as with our robustness checks in section 2, the only way that we have found to obtain the authors' results is to assume that the first documented appearance of moralizing gods is in fact the same time as their emergence; more reasonable alternative approaches yield opposite results.

References

1. Whitehouse, H. *et al*. Complex societies precede moralizing gods throughout world history. *Nature* **568**, 226 (2019).
2. Handy, E. S. C. *Polynesian religion*. (The Museum, 1927).
3. Hogbin, H. I. *Law and order in Polynesia: A study of primitive legal institutions*. (Christophers, 1934).
4. Williamson, R. W. *Religion and social organization in Central Polynesia*. (Cambridge University Press, 1937).
5. Wright, R. *The Evolution of God*. (Little Brown and Company, 2009).
6. Swanson, G. E. *The birth of the gods*. (University of Michigan Press, 1964).
7. Roth, M.T. Law Collections from Mesopotamia and Asia Minor. *Writings from the Ancient World*, no. 6. (Atlanta, Ga. Scholars Press, 1995).
8. Savage, P. E., Whitehouse, H., François, P., Currie, T. E., Feeney, K., Cioni, E., Purcell, R., Ross, R. M., Larson, J., Baines, J., Haar, B. ter, Covey, A., & Turchin, P. (2019). *Reply to Beheim et al.: Reanalyses confirm robustness of original analyses* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/xjryt
9. Whitehouse, H., Savage, P. E., Turchin, P. & Francois, P. Big gods came after the rise of civilisations, not before, finds study using huge historical database. *The Conversation.* Available at: http://theconversation.com/big-gods-came-after-the-rise-of-civilisations-not-before-finds-study-using-huge-historical-database-113801 (2019)**.**
10. Bellah, R. N. *Religion in human evolution*. (Harvard University Press, 2011).
11. Evans-Pritchard, E. E. *Theories of Primitive Religion*. (Clarendon Press, 1965).
12. Wallace, A. F. C. *Religion: An anthropological view*. (McGraw-Hill, 1966).
13. Roes, F. L. & Raymond, M. Belief in moralizing gods. *Evolution and Human Behavior* **24**, 126–135 (2003).

14. Boehm, C. A biocultural evolutionary exploration of supernatural sanctioning. in *Evolution of Religion: Studies, Theories, and Critiques* (eds. Bulbulia, J., Sosis, R., Harris, Erica, Genet, R & Wyman, K.) 143–152 (Collins Foundation Press, 2008).

15. Watts, J. *et al.* Pulotu: Database of Austronesian Supernatural Beliefs and Practices. *PLOS ONE* **10**, e0136783 (2015).

16. Brown, J. S. A Comparative Study of Deviations from Sexual Mores. *Am. Sociol. Rev.* **17**, 135–146 (1952).

17. Purzycki, B. G. Tyvan cher eezi and the socioecological constraints of supernatural agents' minds. *Religion, Brain & Behavior* **1**, 31–45 (2011).

18. Purzycki, B. G. The minds of gods: A comparative study of supernatural agency. *Cognition* **129**, 163–179 (2013).

19. Dobrizhoffer, M. *An Account of The Abipones, An Equestrian People Of Paraguay: Volume 2*. (John Murray, 1822).

20. Turchin, P. *et al.* Seshat: The Global History Databank. Available at: http://seshatdatabank.info/ (2015).

21. Schwimmer, E. G. *Exchange In The Social Structure Of The Orokaiva: Traditional And Emergent Ideologies In The Northern District Of Papua*. (C. Hurst, 1973).

22. Atkinson, Q. D. & Bourrat, P. Beliefs about God, the afterlife and morality support the role of supernatural policing in human cooperation. *Evol. Hum. Behav.* **32**, 41–49 (2011).

23. Hartland, E. S. The 'High Gods' of Australia. *Folklore* **9**, 290–329 (1898).

24. Johnson, D. D. P. God's punishment and public goods: A test of the supernatural punishment hypothesis in 186 world cultures. *Hum. Nat.* **16**, 410–446 (2005).

25. Johnson, D. D. P. Big Gods, small wonder: supernatural punishment strikes back. *Relig. Brain Behav.* **0**, 25–33 (2014).

26. Johnson, D. *God Is Watching You: How the Fear of God Makes Us Human*. (Oxford University Press, 2015).

27. Johnson, D. D. P. The wrath of the academics: criticisms, applications, and extensions of the supernatural punishment hypothesis. *Relig. Brain Behav.* **0**, 1–31 (2017).

28. Lovins, C. Shangdi Is Watching You: Tasan and Big Moralizing Gods. *J. Am. Acad. Relig.* **83**, 464–489 (2015).

29. Nichols, R. & Logan, C. High Gods, Low Gods, and Morality in Ancient China: Developing New Methods, Answering Old Questions. In *Religious Cognition in China: "Homo Religiosus" and the Dragon* (eds. Hornbeck, R. G., Barrett, J. L. & Kang, M.) 161–177 (Springer International Publishing, 2017). doi:10.1007/978-3-319-62954-4_11

30. Norenzayan, A. *Big gods: How religion transformed cooperation and conflict*. (Princeton University Press, 2013).

31. Stark, R. Gods, rituals, and the moral order. *J. Sci. Study Relig.* **40**, 619–636 (2001).

32. Watts, J. *et al.* Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proc. R. Soc. Lond. B Biol. Sci.* **282**, 20142556 (2015).

33. Norenzayan, A. *et al.* The cultural evolution of prosocial religions. *Behav. Brain Sci.* **39**, 1–86 (2016).

34. Steinkeller, P. *Early Semitic Literature and Third Millennium Seals with Mythological Motifs*. (Firenze: Dipartimento di linguistica, Universitá di Firenze. 1992).

35. Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models*. (Cambridge University Press, 2007). doi:10.1017/CBO9781107415324.004

36. Curran, P. J., Obeidat, K. & Losardo, D. Twelve frequently asked questions about growth curve modeling. *J. Cogn. Dev.* **11**, 121–136 (2010).

37. Mirman, D. *Growth curve analysis and visualization using R*. (Chapman and Hall/CRC Press, 2014).

38. Smithson, M. & Verkuilen, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol. Methods* **11**, 54–71 (2006).

39. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2019).

40. Hartig, F. Package 'DHARMa'. (2016).

41. Wickham, H., François, R., Henry, L. & Müller, K. *dplyr: A Grammar of Data Manipulation*. (2019).

42. Skaug, H., Fournier, D., Bolker, B., Magnusson, A. & Nielsen, A. Generalized Linear Mixed Models using 'AD Model Builder'. (2016).

43. Brooks, M. E. *et al.* glmmTMB Balances Speed and Flexibility among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R J.* **9**, 378–400 (2017).

44. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

45. Wickham, H. Reshaping data with the reshape package. *J. Stat. Softw.* **21**, (2007).

46. Wickham, H. *ggplot2: Elegant graphics for data analysis*. (Springer-Verlag New York, 2009).

47. Little, R. J. A. & Rubin, D. B. *Statistical Analysis with Missing Data*. (John Wiley & Sons, 2014).

48. Turchin, P. *et al.* Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *PNAS* **115**, E144-51 (2018).

49. Edward Slingerland et al., in press. "Historians Respond to Whitehouse et al. (2019), 'Complex Societies Precede Moralizing Gods Throughout World History,'" *Journal of Cognitive Historiography.*

50. Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.18.2. http://mc-stan.org/.

51. Gelman, A. & Tuerlinckx, F. Type S error rates for classical and Bayesian single and multiple comparison procedures. *Comput. Stat.* **15**, 373–390 (2000).