Big Gods and big science: further reflections on theory, data, and analysis

Peter Turchin, Harvey Whitehouse, Jennifer Larson, Enrico Cioni, Jenny Reddish, Daniel Hoyer, Patrick E. Savage, R. Alan Covey, John Baines, Mark Altaweel, Eugene Anderson, Peter Bol, Eva Brandl, David M. Carballo, Gary Feinman, Andrey Korotayev, Nikolay Kradin, Jill D. Levine, Selin E. Nugent, Andrea Squitieri, Vesna Wallace & Pieter François

Routledge
Taylor & Francis Group

RESPONSE

🔓 OPEN ACCESS   Check for updates

Big Gods and big science: further reflections on theory, data, and analysis

Peter Turchin [a,b,c], Harvey Whitehouse [c], Jennifer Larson[c,d], Enrico Cioni[e], Jenny Reddish[a], Daniel Hoyer [e,f,g], Patrick E. Savage[h], R. Alan Covey [i], John Baines [j], Mark Altaweel[k], Eugene Anderson[l], Peter Bol[m], Eva Brandl[n], David M. Carballo[o], Gary Feinman[p], Andrey Korotayev [q], Nikolay Kradin[r], Jill D. Levine[e], Selin E. Nugent[s], Andrea Squitieri [t], Vesna Wallace[u], and Pieter François[c]

[a]Complexity Science Hub, Vienna, Austria; [b]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA; [c]Centre for the Study of Social Cohesion, University of Oxford, Oxford UK; [d]Department of Modern & Classical Languages, Kent State University, Kent, OH, USA; [e]Seshat: Global History Databank, Evolution Institute, San Antonio, FL, USA; [f]George Brown College, Toronto, Canada; [g]Evolution Institute, San Antonio, FL, USA; [h]Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan; [i]Department of Anthropology, University of Texas at Austin, Austin, TX, USA; [j]Faculty of Oriental Studies, University of Oxford, Oxford, UK; [k]Institute of Archaeology, University College London, London, UK; [l]Department of Anthropology, University of California Riverside, Riverside, CA, USA; [m]Department of East Asian Languages and Civilizations, Harvard University, Cambridge, MA, USA; [n]Department of Anthropology, University College London, London, UK; [o]Department of Anthropology and Archaeology Program, Boston University, Boston, MA, USA; [P]Negaunee Integrative Research Center, Field Museum of Natural History, Chicago, IL, USA; [q]Higher School of Economics, National Research University, Moscow, Russia; [r]Institute of History, Archaeology and Ethnology, Vladivostok, Russia; [s]Institute for Ethical AI, Oxford Brookes University, Oxford, UK; [t]Historisches Seminar, Ludwig-Maximilians-Universität München, Munich, Germany; [u]Department of Religious Studies, University of California Santa Barbara, Santa Barbara, CA, USA

Our target article empirically tested the Big Gods Hypothesis which proposes that beliefs in moralizing supernatural punishment (MSP) contributed to the evolution of socio-political complexity (SPC) in world history. We tested this hypothesis using a suite of measures of MSP, SPC, and other potential evolutionary drivers coded in Seshat: Global History Databank. Our analyses indicate that intensity of warfare and productivity of agriculture were major drivers in the evolution of both SPC and MSP. The correlation between social complexity and moralizing religion resulted from shared evolutionary drivers, rather than from direct causal relationships between these two variables. Most commentaries on the target article broadly accept our conclusions, but some argue that alternative measures might be used in future studies before the Big Gods Hypothesis can be conclusively rejected. In this response, we argue that while some of these alternative measures should be developed, they are closely related to the ones we have already adopted. Thus, it seems unlikely that further research will give rise to substantially different outcomes. A particularly fruitful aspect of the discussion is that it illustrates both the pitfalls and productive affordances of transdisciplinary research that seeks to bridge the "two cultures" of the humanities and sciences.

Our target article has given rise to some thought-provoking commentaries and we welcome the opportunity to respond. The points raised can be conveniently grouped under four main headings, dealing with issues of theory, data, analysis, and future directions. Some particularly valuable points have been made in several commentaries about the theoretical foundations of the Big Gods Hypothesis and more broadly the causes and consequences of changes in religious belief as societies have

grown in size and complexity. As several commentators make clear (e.g., Lawson; Geertz; Schlieter & Rota), and our target article also shows, MSP was likely a consequence rather than a cause of fundamental changes in society driven by agriculture and warfare but operating in complex ways at multiple levels. But commentaries also raise broader questions about whether more elaborated forms of puritanical zeal (Fitouchi et al.) and transcendentalist thinking (Strathern) might have impacted the way beliefs in MSP affected cooperation in increasingly large and complex societies.

The productiveness of these theoretical debates, however, depends in turn on the quality of the data used to test competing hypotheses. Some of the commentaries raise valuable questions in that regard, for example concerning the extent to which agricultural variables (Kaše & Glomb) can be used as proxies for affluence. Purzycki et al. also critique our data, but unfortunately did not appear to consult the 210 pages of detailed Analytical Narratives (http://seshatdatabank.info/databrowser/moralizing-supernatural-punishment-narratives.html) or the 219 pages of tables justifying our inferences of presence/absence of MSP based on these narratives (http://seshatdatabank.info/databrowser/moralizing-supernatural-punishment-nga_tables.html) despite their being clearly linked in the "Data Availability and Supplementary Online Materials" section of the target article. If they had, they could not have come to the mistaken conclusion that our codings of MSP as "Inferred Absent" were not "based on the ethnographic, historical, and archaeological records."

We also consider various challenges posed by transdisciplinary research that combines strategies of interpretive analysis typical of humanities disciplines and statistical analysis typical of big data science. Researchers working on one side of the divide struggle to understand the methods and findings of those on the other. Much closer dialogue and collaboration is required to bridge the silos. In this subsection of our response, we also discuss some major issues of concern with regard to the statistical analysis conducted by Purzycki et al. We show that these critics have fundamentally misunderstood the conceptual and statistical approach used in the target article. Furthermore, their simulation model, which, according to them, demonstrates that our approach cannot test evolutionary causality, makes deeply flawed assumptions about the process generating data, with the result that their simulation results are irrelevant to the issues at hand.

Finally, we consider the implications of the above discussion for the future of research and debate on the Big Gods Hypothesis. Although there is scope to explore further evidence in future studies, we argue that our latest analyses of Seshat data, and the resulting response from the scientific community, have greatly diminished the plausibility of the Big Gods Hypothesis as originally formulated. But these discussions also open up interesting avenues for future enquiry.

Theory

Seshat: Global History Databank was created to test theories about cultural macroevolution, with a particular focus on the evolution of Socio-Political Complexity (SPC). In this framework, SPC can serve as a predictor variable (causing various other features of culture to evolve) or as outcome variable (the consequence of various novel cultural innovations). One of the greatest challenges for the Seshat project has been to decide which theories using SPC as either predictor or outcome variable should be prioritized for testing. In his commentary, Lawson argues that priority should be given to testing hypotheses well-grounded in theory and evidence from the cognitive and evolutionary sciences. Lawson points to an influential study by Baumard & Boyer (2013) as a model example on the grounds that it draws on life history theory. The idea that affluence played a role in MSP is plausible (see also Baumard et al., 2015), which is why we included it in our predictor variables. However, while life history theory is well-grounded in evolutionary biology, the way this concept is applied in psychology is very different (see Nettle & Frankenhuis, 2020) and there is scope for debate about its applicability in the present context. This is also the spirit in which we decided to prioritize the Big Gods Hypothesis for testing. As we explain in our target article, the earliest formulations of this hypothesis held that MSP drove the initial rise of SPC (Norenzayan, 2013) while later formulations argued that MSP was among a suite of other variables that helped SPC to

consolidate and grow once it had already become established (Norenzayan et al., 2016). Both versions of the hypothesis seemed promising because they were grounded in extensive empirical research using psychological experiments and surveys based on well fleshed-out theoretical ideas, even though the literature is fraught with conceptual and methodological problems, lacunae, and contradictory findings (see McKay & Whitehouse, 2014). On the question of whether or not the Big Gods Hypothesis would be supported by the analysis of global historical data, however, we remained entirely agnostic prior to our investigations. But, as we explain in the target article and in the discussion below, the overall weight of evidence does not support the Big Gods Hypothesis. One of the most important scientific functions of the quantitative analysis of world history using databases like Seshat is precisely to winnow out theories that make seemingly plausible and partially supported claims but which turn out to be wrong.

That said, perhaps the Big Gods Hypothesis would find more support if it were modified in some way to better capture the contribution of increasingly well-developed moralizing religions to large-scale cooperation. This appears to be the spirit in which Fitouchi et al. argue that the theory of Big Gods should be expanded to include a special focus on the characteristics of what they call "puritanical religions."

While Fitouchi et al. acknowledge that our multi-dimensional MSP measure is an advance on binary ones, they argue that a missing but important dimension is what they describe as puritanical zeal—a quality they maintain is especially effective in quelling selfish impulses. More specifically, they propose a spectrum of ideal types of moralizing religion, of which puritanical expressions constitute the most developed form. The least developed, which they assert is commonplace in small-scale societies, is belief in supernatural punishment of a limited range of transgressions. And somewhere between these two ends of the spectrum, they argue, are "big gods," with broad moral jurisdiction, but not puritanical elements. They argue that this in turn would require an expansion of the Seshat codebook.

According to Fitouchi et al., puritanical zeal finds various forms of expression in a number of contemporary religions (they mention specifically Christianity, Confucianism, Hinduism, Islam, Buddhism) and some ancient ones (they mention Platonism and Stoicism). Puritanism does not just proscribe immoral behaviors (such as stealing and unauthorized killing), but also postulates supernatural punishment for behaviors and mental states that only contribute indirectly to immoral conduct. Proponents of this view argue that cleansing one's mind of transgressive thoughts and feelings could be particularly effective at promoting prosocial conduct. However, this would imply that mental discipline is more important for cooperation than the enforcement of norms on the ground —for example by formal legal codes, punishment in courts, and armed security in towns such as night watches. This is undoubtedly an intriguing possibility, although the theoretical rationale and evidential basis would benefit from further elaboration.

For the purposes of our target article, we captured (among other measures of MSP) at least two of the three dimensions that Fitouchi et al. regard as important. Thus, we included a measure of whether supernatural punishment was "targeted"—that is, focused specifically on culpable individuals. And we also included a measure of whether it was "broad"—that is, encompassing a wide range of moral transgressions in human affairs, equivalent to the midway level envisaged by Fitouchi et al. What the latter would argue we did not include, and perhaps should have done, was a measure of puritanical zeal (as conceptualized above). The question is whether our analysis would have been strengthened by doing so.

The answer to this question hinges crucially on whether puritanism might have made a significant contribution to the evolution of socio-political complexity. Recall that our primary aim in the article was to test the so-called Big Gods Hypothesis which proposes that beliefs in MSP contributed to that evolutionary process. Since the Big Gods Hypothesis has itself undergone significant changes (as noted in the target article), we attempted to capture variables capable of testing both earlier and later versions. But we went further still, by adding a measure of the overall primacy of moralizing concern, which would certainly include (but not be limited to) expressions of puritanical zeal, as conceptualized by Fitouchi et al.
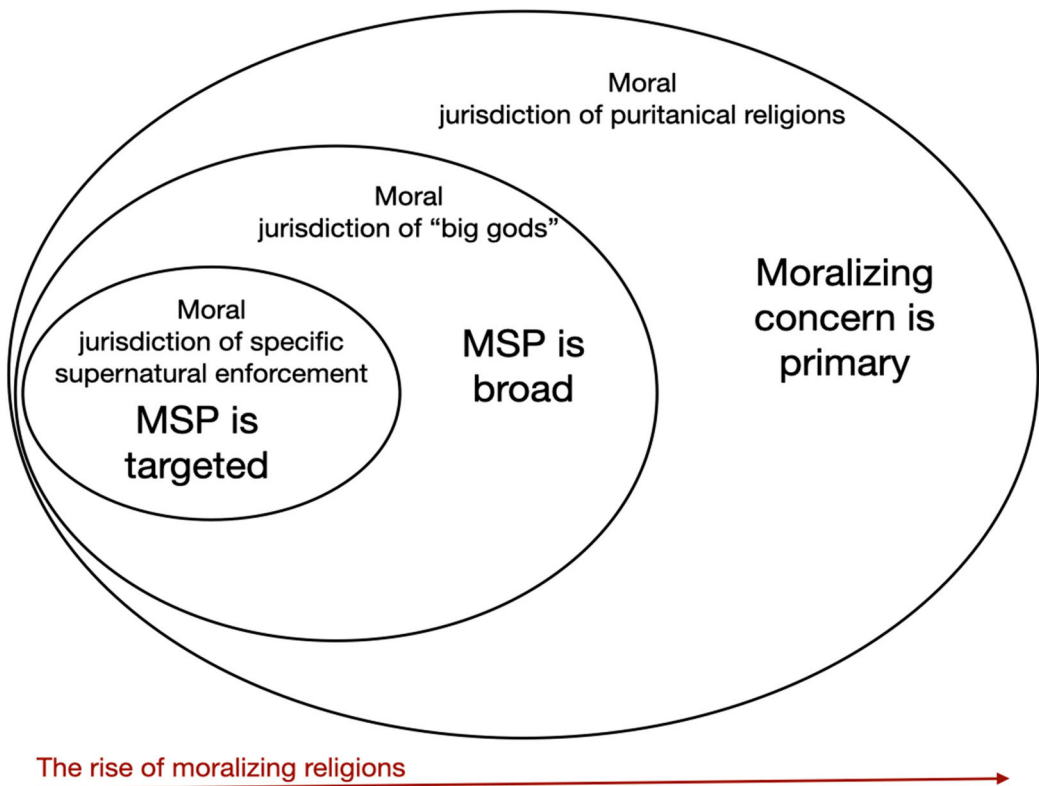
Figure 1. Fitouchi et al.'s scheme revised using MSP measures from the target article.

We would argue that our measure of the primacy of moralizing concern is effective in capturing a pervasive focus in the religious system on the enforcement of morality in human affairs, as distinct from the less exacting requirements of MSP that is just targeted or even broad but not the primary concern of the tradition. Indeed, the diagram presented by Fitouchi et al. could be readily adapted to encompass these measures that we actually used in our target article (see Figure 1).

Presumably, Fitouchi et al. would argue that our measure of moralizing concern as primary is not sufficient to capture puritanism. But, if so, what evidence is there that puritanism deserves special attention in the context of testing the Big Gods Hypothesis (or variants of it)? While there may be some intuitive appeal in the idea that "inner moral reform" could increase levels of cooperation relevant to large-scale social formations, the evidence for this is currently scant. The psychometric measures that come closest to capturing interiorized morality are perhaps those used in the moral cleansing literature which seeks to assess the effects of transgressive thoughts and behaviors on one's perceived moral balance sheet (e.g., Jordan et al., 2011; Ketelaar & Au, 2003; Lee & Schwarz, 2011; Lee & Schwarz, 2021; Sachdeva et al., 2009; but cf. Ross et al., 2021). Presumably, puritanical religions not only increase the salience of internal mental states that have moral consequences, but also commonly provide ritualistic methods of achieving absolution, or at least some kind of catharsis. Priming studies suggest that acts of purification (e.g., handwashing) may actually reduce any potential benefits that puritanism may have for cooperation by restoring the moral balance sheet of transgressors, freeing them to sin all over again (Zhong & Liljenquist, 2006) There is some interesting evidence that ritualized confession increases prosocial behavior (such as donations) towards religious specialists, perhaps out of gratitude to those granting absolution (McKay et al., 2013) although this is not in itself sufficient reason to think that MSP measures that fail to include puritanical zeal are inadequate. We agree

that an interesting avenue for future research may be to expand our measure of MSP to include variables such as "zeal" and the rituals that enforce it, particularly exploring whether additional developments in moralizing religion are related to continued increases in SPC once these beliefs have already become well established.

Other commentators on our target article were more inclined to accept that the measures we used were adequate for the task of testing the Big Gods Hypothesis and are broadly persuaded by our conclusions. Geertz, for example, notes that our main findings are very much in line with those that he and others had already anticipated based on more qualitative considerations (e.g., Geertz, 2014; Stausberg, 2014). Geertz points to numerous other factors that likely contributed to the scaling up of cooperation over the course of world history including intergroup conflict, trade, agriculture, writing, and the spread of disease. As he correctly observes, the Seshat project has sought to systematize many of those factors and include them in the analysis here or in the analyses conducted in other papers using Seshat data (e.g., Turchin et al., 2022).

An important implication of Geertz's reflections on these issues is that the causal relationships between MSP and other variables captured in our analyses are complex and may operate, as Geertz also explicitly notes, at multiple levels. It is important to disambiguate the levels at which different kinds of causal relationships come into play in cultural and socio-political evolution. Moreover, the evolutionary framework adopted by the Seshat project is guided by cultural multilevel selection (CMLS) theory, which incorporates such evolutionary mechanisms as inter-polity competition and adaptive borrowing of successful traits.

The conclusions of our target article, of course, go well beyond the negative claim that MSP does not drive the rise of SPC and includes a set of novel hypotheses that are better supported by our analyses. An example is our claim that mounted warfare drove the evolution of a number of other cultural domains, including MSP. Schlieter and Rota (2022) make the important point, however, that if warfare drives MSP then we need to explain why this occurred only in some cases and not others. This takes us into some very interesting theoretical territory, explored more fully in other Seshat publications.

For example, several scholars have proposed that MSP is part of a much broader cluster of features associated with the so-called Axial Age (Mullins et al., 2018). For example, the rise of Confucianism and the other "Hundred Schools of Thought" during the late first millennium BCE in North China carried not only a moralizing force but were attended by universalizing concepts and espoused some egalitarian ideals; likewise, the political and legal reforms of Classical Greece developed hand-in-hand with the spread of Platonism and off-shoot philosophies (Levine & Haar, 2019; Reddish & De Angelis, 2019). At the same time, it is unclear exactly how strong this pattern is, or the extent that these different features have coevolved in different places, a question requiring further scrutiny (Whitehouse et al., 2019). While classic theorists have proposed that "axiality" emerged in the first millennium BCE, we found that in some regions of the world (e.g., Egypt) these traits already emerged much earlier (Reddish & Manning, 2019). Axiality may be best characterized not so much as an "age" but a "stage" in the rise of SPC (Whitehouse et al., 2019) and, if so, this would apply as much to MSP as to the other features associated with axial religions, ethics, and political philosophies; though the exact correspondence between these different developments requires further elaboration, the subject of ongoing research. The defining feature of this stage in the rise of SPC is the presence of multiple competing interest groups and identities resulting from trade, conquest, and immigration. As multi-ethnic empires grew more internally heterogenous, extreme forms of inequality and top-down coercion would have stoked rebellion and internal conflict. Axial principles provided solutions to this problem, arguably allowing early adopters to win out against empires that were slow to appreciate the benefits that these principles conferred (Bellah, 2011). To the extent that mounted warfare commonly preceded the adoption of MSP and other Axial features, it makes sense that the former helped to spread the latter and that both in turn contributed to further increases in SPC. Further, as we've shown elsewhere (Turchin et al., 2021), the spread of mounted warfare and the intense inter-state pressure this carried

impacted certain regions—largely those neighboring the central Eurasian steppe—earlier than others. This may help answer Schlieter and Rota's question about why we see these dynamics more prominently, or at least earlier, in some areas than others. These conclusions are also broadly consistent with Strathern's points about the growth of transcendental religions, although they do not resolve his question about the relative importance of inventing vs. borrowing axial features such as MSP.

Data

As important as the grand theoretical debates discussed in the last section undoubtedly are, a test of competing theories is only as good as the data on which it is based. Testing theories using historical data requires methods that avoid cherry picking (selection bias) while also capturing variables of theoretical interest that serve as adequate proxies. Cherry picking is an inevitable problem when theories about history are tested in a purely anecdotal way, by choosing evidence that supports the theory while ignoring evidence that does not. Even when attempting to be even-handed by choosing a mix of examples that support and refute a theory, we are generally powerless to establish the relative weight of the arguments for and against. The only solution to such problems is to quantify history, but this too raises many challenges, ranging from autocorrelation (the nonindependence of data points) to the practical problem of how to organize the data in a form than can be subjected to quantitative analysis. We say more about the latter issue shortly. But first, we turn to an even more basic problem which is whether the variables necessary to test theories about the role of MSP in the evolution of SPC can be adequately captured in Seshat data. Kaše and Glomb argue that this issue is present in our test of Baumard & Boyer's hypothesis about the role of affluence in the rise of MSP, using agricultural measures as proxies for affluence.

Kaše & Glomb question the adequacy of agriculture as a proxy for affluence, when testing the Baumard hypothesis. We argue that our use of agricultural productivity passes the test in that it has both solid theoretical grounding as well as empirical validation for being one (among many) measures of general affluence. Still, the critique is well made and we agree that future work is necessary to fully flesh out Baumard and colleagues' affluence hypothesis. Better use of archaeological indicators of affluence, such as settlement density, typical house size, or relative abundance of epigraphic material as Kaše & Glomb propose would add much needed clarity to these relationships. As noted in the target article, we employ here the best empirical data we had available, while realizing that more and better proxies will be developed in the future.

One of the consequences of our coding scheme for MSP data is that it produced codes of "absent" or "inferred absent". Fitouchi et al. appear not to have understood this, for example where they argue that the "baseline" of MSP is non-zero and thus all societies have some degree of MSP. However, according to the definitions used in our target article, many "primary religions" (see Assmann, 2010) appear to involve no MSP (moralizing supernatural punishment and reward). Following Theo Sundermaier, Assmann distinguishes between "primary" and "secondary" religions. Primary religions are those which "evolve historically over hundreds and thousands of years within a single culture, society and generally also language, with all of which they are inextricably entwined." Examples include the Eurasian polytheisms (e.g., Mesopotamian, Greek, Norse), the indigenous religions of the American civilizations (e.g., Maya, Inca, Aztec) and many tribal religions. Secondary religions are those "that owe their existence to an act of revelation and foundation." These traditions react to primary religions, denouncing them as paganism, idolatry or superstition (Assmann, 2010, p. 1). Secondary religions are "counterreligions" which (at least initially) adopt an oppositional stance, and in keeping with their claims of revelation, they are also typically religions of the book. Unlike primary religions, they are translocal and (ex)portable. We go on to use the word "primary" in a different sense—referring to our coding scheme and whether moralizing is the primary concern of the religious system. Of the primary religions found in the natural geographic areas on which our analysis is based, nine are considered by our experts to

have no MSP, 16 limited MSP, two have high MSP, in four cases there is a near-complete lack of information or too much scholarly disagreement to make a judgment, and two (Kansai and South China Hills) were excluded due to lack of expert approval. The most notable geographical pattern is the apparent absence or very limited presence of MSP in several primary religions of the Americas. Of the seven primary American religions in our sample, our experts indicated that MSP was minimal or absent in five, and there is scholarly disagreement about its presence in the other two.

Secondly, the additional measure proposed by Fitouchi et al. is more moralizing even than those in which moral concern is primary (our maximal measure of MSP in the target article). To clarify, the target article's measure of "primary" asks whether interpersonal morality is the primary concern in the religious system, and our terminology classifies as "moralizing religions" only those that meet this high standard (rather than any religion with at least some MSP). We agree that moralizing religions may vary in the degree to which beliefs and practices focus on an individual's inner disposition and self-discipline. However, the relationship between MSP and "puritanical" or "disciplinary" religions as described by Fitouchi et al. is not clear-cut. While Confucianism is much concerned with self-discipline and ethics, for example, we have found little or no unambiguous evidence for MSP targeted toward the individual's failure to meet Confucian standards of propriety. Likewise, the role of MSP in Greek philosophical movements is unclear. Pythagoreanism, for example, was deeply concerned with ritual purity; while the Pythagorean *akousmata* include some clearly moralizing precepts, they are dominated by cultic and dietary rules (Diogenes Laertius, 1925, 8, 17-20; Thom, 2020), such that fate in the afterlife is tied to ritual requirements at least as much as interpersonal ethics. In their pursuit of purity, movements like Pythagoreanism often focus obsessively on sexual morality, dietary restrictions and other forms of renunciation of the physical world, which in themselves help the practitioner to transcend the material (Parker, 1983, pp. 294–299). Ritual purity rules are not necessarily moralizing in ways that contribute to cooperation and social complexity, the subject of our paper, but elaborate evolved dispositions toward hazard precaution (Boyer & Lienard, 2007) and the emotion of disgust (Kazen, 2014). While the Pythagoreans undoubtedly believed in reincarnation, recent scholarship argues that Pythagorean metempsychosis was non-retributive, and the early sources on Pythagoreanism do not mention supernatural punishment (Pellò, 2018; Zhmud, 2012, p. 230). The ethical system of the Stoics, meanwhile, was highly developed and focused on self-discipline in ways that encouraged social cooperation, yet MSP is all but absent in the major ancient account of Stoic ethics (Diogenes Laertius, 1925, 7, 83–89) and Stoic philosophers disagreed on the role of punishments in divine providence (van Houte, 2022). We agree with Fitouchi et al. that broad, continuous measures are key to unpacking the different formulations of ritual adherence, ethical purity, supernatural and human punishment, and other categories that comprise moralizing religions around the world. We look forward to seeing further tests of the relationships between these various factors, SPC, inter-state conflict, and any number of other critical elements.

Geertz raises a separate critique of our data quality, referring to a recent critique leveled by Franziska Naether (2021). As Geertz suggests, the proper venue to discuss these misunderstandings is in our "future responses" to the commentaries by Naether and others in that particular issue. We plan to respond to these points in a separate paper but, in essence, most of the specific criticisms are based on methodological misunderstandings and do not warrant the negative conclusions drawn.

While we welcome the valuable discussions opened up by the suggestions from Kaše & Glomb regarding our use of agricultural variables as proxies for affluence and from Fitouchi et al. regarding the scope of our MSP variables, as well as the many useful observations put forward by Geertz, Purzycki's et al.'s critique appears to be based on a misunderstanding. They assert that "the Seshat database has replaced the bulk of its missing values of moralistic supernatural punishment with 'inferred absences'" without any basis on "ethnographic, historical, and archaeological records." This is incorrect—Purzycki et al. appear to have failed to consult the 210-page document of Analytical Narratives we went to great lengths to produce precisely in order to correct our previous error by which

we failed to accurately distinguish between cases where scholarly consensus on MSP was truly "unknown" vs. those where there was ample evidence with which to infer their absence. This error was clearly described in our short note retracting our previous paper in response to a critique coauthored by Purzycki (Beheim et al., 2019): "we accept that we should have labeled moralizing gods as 'absent' or 'inferred absent' rather than 'unknown' in portions of our dataset before the dates of the first appearance, rather than converting 'NAs' to zeros during the phase of analysis" (Whitehouse et al., 2021). Since we recognized this error, we have gone to great lengths to undertake a process of data review in which we have ensured that the data are now correctly labeled. We also provide extensive information based on scholarly sources to buttress those labels. Details of how that process was undertaken are provided at length in the target article so there is no excuse for stating that we have simply replaced missing values with inferred absences without justification.

Examining our Analytical Narratives makes clear the errors in Purzycki et al.'s claims. For example, they state: "there does not appear to be a single written record in the Seshat database that documents moralistic gods among small-scale societies prior to ethnography." Further, they say, "the moralistic traditions are overwhelmingly Christian, Buddhist, Hindu, and Muslim where copious amounts of documentation" are available. They thus imply that the Seshat database is biased against the possibility of MSP in small-scale societies or in cultural contexts where writing is absent. This is far from accurate. In fact, relying on the opinions of our expert consultants, we coded for the presence or inferred presence of some degree of MSP in Latium (3600 BCE-265 BCE), Paris Basin (3200 BCE-27 BCE), Iceland (930 CE-1000 CE), Sogdiana (2000 BCE-751 BCE), among the Akan people of the Ghanaian Coast (1501 CE-1894 CE), among the Sakha of the Lena River Valley (1400 CE-1900 CE), among the Iban of the Kapuas Basin (1650 CE-1987 CE) and on Big Island Hawai'i (1000 CE-1819 CE). In the early part of the listed periodizations, all of these were small-scale societies with minimal or no use of writing, and practiced indigenous/primary religions. We also acknowledge scholarly disagreement on the presence of MSP before ethnography among the Tairona of North Colombia (1050 CE-1524 CE) and among the Aztecs (400 BCE-1519 CE).

Purzycki et al. complain that we do not recognize moral concern as primary in religions that are not Christian, Buddhist, Hindu, or Muslim. In fact, we infer that moralizing concern had become primary in Egyptian religion by the Old Kingdom (third millennium BCE) and that it was primary in the indigenous religion of the Chuukese. Where evidence exists to support such an inference, we are demonstrably willing to recognize it. However, the study of world history demonstrates that what Alan Strathern calls "immanentist" religions, the primary mode of religion of early and small-scale social formations, are less likely to be highly moralizing. In immanentist systems, morality is situational and communal. While some aspects of the supernatural may be connected to prosocial norms in these societies (such as ideas about loyalty to kin), these religions tend to lack abstract, universalizing ethical codes. Internalizing systematized ethical precepts is typically not at the center of religious life, and the latter focuses instead on channeling supernatural power to benefit the community (Strathern, 2019). We agree that moralizing supernatural punishment exists in some small-scale societies, and our data clearly documents its presence; it simply does not change our conclusions. Nor do the sources cited by Purzycki et al. demonstrate the point they wish to make, that small-scale societies have high levels of MSP. In Boehm (2008), for example, the sample of 18 "tribal" cultures shows that "nonmoral taboos" are subject to supernatural punishment more consistently than interpersonal transgressions, a pattern typical of immanentist traditions.

As to our use of the code INFERRED ABSENT, each instance is supported by a written rationale in the coding tables. All coding decisions were made by the Data Review Board (defined in the main article) through consultation with a domain expert. For the Paris Basin in the period 3200 BCE-27 BCE, for example, we code INFERRED ABSENT for the variable "Moral concern is primary" in the religious system, writing "No known polytheistic religion of the pre-Christian era in Europe displays primary concern for interpersonal morality. We infer similar beliefs backwards as far back

as the Bell Beaker period." This does not mean that all MSP is inferred absent. For the same period, we code INFERRED PRESENT for the variable "Moralizing enforcement is targeted," noting that "Comparisons of oath formulas across a broad range of Indo-European [language] families suggest that belief in MSP against oath-breakers has deep roots in Indo-European culture." We do not infer that MSP is absent in cases where information is simply lacking. For example, we code UNKNOWN for MSP among the Mande people of the Inland Niger Delta, given that the archaeological and ethnographic evidence for this culture is insufficient to infer either the presence or absence of MSP. Our codes, thus, contain all of the diversity and sensitivity to the archaeological record that Purzycki et al. suggest we overlook; we recognize aspects of MSP in small-scale societies, we differentiate between what is known and inferred, and we allow for areas that are simply not knowable based on the present state of research.

Analysis

Some of the commentaries voice concerns about the difficulty of evaluating research that mixes traditional methods of interpretative analysis typical of humanities disciplines with the statistical methods of analysis typical of big data science, given the rarity of individuals who are experts in both forms of data analysis. The problem is raised in the commentaries from the perspective of humanities scholars. As Strathern puts it concisely: "I am unable to interrogate the core methodology deployed here, the use of regression analyses, and the crucial question of how sociopolitical complexity may be associated with MSP without being identified as a cause of it." And as Geertz further elaborates: "How do we gain insight into competing claims based on mathematically opaque arguments? … I, for one, am mystified, and not a little frustrated." An analogous problem, however, is also apparent the other way around, from the perspective of the data scientist. As noted in the preceding section, Purzycki et al. do not appear to engage with our data, dismissing much of it as simply missing. But whereas humanities scholars are inclined to acknowledge their inability to assess statistical claims, data scientists are prone to overstating their knowledge of history and archaeology. Such proclamations are confounding to many humanities scholars whose careers are devoted to interpreting evidence from archaeological, ethnographic, and archival sources pertaining to societies without writing.

Such difficulties are inherent in any massively transdisciplinary project, such as Seshat. We would argue that the best way to overcome them is to work in teams that span multiple disciplines, from humanities to social sciences to data science. This is the approach that the Seshat project has adopted. We recognize that this is indeed "frustrating" to any individual with understandably discipline-specific expertise, but we believe it is the most effective way to address big questions about the evolution of social complexity. A key dimension of our approach is that we aim to satisfy not only a diverse set of scientists who have produced this research (as part of the Seshat group), but also external evaluators of the reported results.

For this reason, the Seshat project, from the beginning, has been committed to open-sourcing both the data and the analysis methods. We aim to ensure that the data are as rigorously sourced as the relevant humanities disciplines will allow, and that the analyses employ the most current, relevant techniques favored by data scientists. Humanities scholars are invited to critically examine the decisions underlying coded data by consulting the data tables and the more extensive, more qualitative analytic narratives. Similarly, data analysts are provided with Supplementary Materials that include data files and analysis scripts, so they can examine the logic of the statistical analysis, replicate its results, and potentially offer improved solutions. This team-based approach requires mutual trust: humanities scholars will have to rely on an independent assessment by data scientists of the statistical methods, while data scientists need to trust the judgment of historians and archaeologists underlying the data.

Disciplines such as history, archaeology, and philology have developed sophisticated and articulated approaches to the historical study of religion. Instead, Purzycki et al. dismiss the expertise of

these scholars as inferior to their own views of what is known, and what is unknown about the human past.

Equally unfortunately, this particular team of scientists also fails to evaluate properly the statistical approaches used in the target article. They know one particular approach to causality, which uses directed acyclic graphs (DAG). However, our statistical approach to causality, as explained in the target article, uses a different conceptual framework, dynamic regressions (DR), which is based on the ideas of Norbert Wiener (1956), later developed by Clive Granger (1969).

The most important difference between DR and DAG is that the latter does not explicitly include the time dimension. Thus, instead of specifying how one event or variable at one point in time affects another variable or event at a later point in time ($X_t \rightarrow Y_{t+1}$), as is done in DR models, in DAG causal connections are denoted without time subscripts ($X \rightarrow Y$). In other words, scenarios of mutual causation cannot be investigated. Furthermore, the main goal in DAG is estimation of the causal effect. This approach is appropriate if we need to know, for example, by how many years a particular drug would increase life expectancy, when we only have observational data. To answer this question, an analyst must assume a particular DAG—its form is under-determined by (time-unresolved) data.

The goal of the DR approach is different, because we aim to use data to adjudicate between different theories of social evolution, each proposing a different causal graph (an example is in Figure 3 of the target article). This is generally impossible to do with static (time-unresolved) data, which is why the goal of the Seshat project from its inception was to collect time-series data. Unlike with DAGs, the main goal of the DR approach is model selection, choosing which predictor terms should be included on the right-hand side in Eqn. 1. We are also interested in estimation, because we want to compare the numerical strengths of different factors, but this goal is secondary to model selection, as we first need to determine which causal graph should be used for coefficient estimation. Thus, the DAG approach, excellent as it is, differs in goals and techniques from the DR approach. The DR approach was designed to resolve questions of causation in evolutionary processes ("descent with modification") that unfold slowly in time. It allows us to deal with such complications as mutual causation loops and temporal autocorrelations arising from the inertial nature of evolution, and (with fairly straightforward extensions, see *Methods* in the target article) with spatial diffusion and phylogenetic effects. A model's ability to predict data is interesting not in itself, but as a tool for adjudicating between different theories.

In short, the dynamic regression analysis attempts to distinguish correlation from causation by estimating what influence potential causal factors at a previous time have on the response variable at a later time. Purzycki et al. seem not to understand this crucial point, which is why they make statements such as "the direction of causality cannot be determined by data (or regression coefficients) alone." This statement is correct for an analysis of *static* data, but not true for time-resolved data. If a theory predicts that a factor X causes Y, but we find that Y precedes X in time, this observation falsifies the theory. Of course, there are complicating factors that require caution in analyzing time-resolved data, which are further discussed in (Turchin et al., 2022) but generally a temporal dimension allows us to approach questions of causality much more effectively than is possible with static data alone.

Purzycki et al.'s misunderstanding is particularly glaring in the simulation model that they use to illustrate their point that our approach cannot yield accurate causal inferences. First, their model has no time dimension, whereas an explicit time dimension is the key to the DR approach, as explained above. Second, they start with a DAG (see their Figure 3) based on the one in McElreath (2020, p. 515). Their R script then implements the DAG in a mechanical way, which actually does not make sense. Thus, "Missingness" is simply a random variable that is positively correlated with "Writing" (although we would expect that Missingness would decrease in the presence of texts). Worse, Missingness then has positive effects on observed Social Complexity and Moral Gods. In other words, high Missingness increases the value of observed Moral Gods. This is nonsense, because what Missingness actually does is *convert* some values of observed Moral Gods into missing

values. Instead, in the Purzycki et al. model, Missingness *grows* Moral Gods. Naturally, the "results" from a model with such inappropriate assumptions can have no bearing on the validity of the DR approach.

In addition to this major misunderstanding relating to the difference between a DAG and a DR approach, which invalidates their critique, Purzycki et al. also got a number of minor points wrong. For example, they criticize us for using a multiplicative scale for aggregating MSP elements into an overall MSP measure and suggest that an additive scheme is better. Yet, the target article clearly states (in the section on *Aggregating Binary Codes into an Overall Measure of Moralizing Religion*), "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)." In *Supplementary Results* we report that using an additive measure of MSP does not change the results.

Lawson, Geertz, Fitouchi et al. and to some extent Strathern make different points relating to appropriate levels of analysis. They stress the potential importance of cognitive processes among individuals (or at least small groups of individuals) in driving the adoption of novel forms of ideological thought, such as increasingly moralizing or transcendentalist religions. These are valid points, though beyond the scope of the present study which seeks to explore these dynamics at the more macro, polity-centred level. Future work is needed to explore the various ways in which more micro-level, individual cognitive structures and social-psychological incentives—likely responding to the larger structural pressures identified here—affect insight into the rise and spread of moralizing religions. We look forward to fruitful collaborations on these and related topics.

Conclusions and future priorities

By social science standards, the Seshat database is a very expensive scientific instrument. It cost millions of pounds to build, requiring vast amounts of expert labor over the past ten years. The codebook was originally designed to establish all-encompassing measures of SPC and of many of the predictor variables that are thought to have contributed to its evolution, including Big Gods. Each time we refine and improve our measures of particular predictor variables, for example by extending our binary measures of moralizing religion to include our current spectrum of MSP measures or coding information about crop productivities in different parts of the world, we have had to invest substantial resources in the process. Whenever we add a new variable to Seshat, it needs to be coded in hundreds of polities, and this is obviously a very costly business. It is, therefore, necessary to be highly selective when deciding which new hypotheses to test.

How are we to prioritize the many useful recommendations in the commentaries on our target article? One way to approach this problem is to start with the empirical needs of particular theories and their associated hypotheses. We think we have given the Big Gods Hypothesis a very good run for its money: first, we sought to date the first appearance of "big gods" using a binary measure; now, we have employed a gradual measure of MSP combined with statistical methods that allow us to explore causality as well as sequence. We find that the Big Gods Hypothesis, in the various forms it has been formulated up to now, is not supported. While it is of course tempting to try to find yet further ways of interpreting the Big Gods Hypothesis, for example by developing measures of puritanical zeal, we have chosen to focus for the foreseeable future on broadening our variables to include other dimensions of religion (e.g., ritual sacrifice) and other domains of culture (cf. Turchin et al., 2022). We do, though, encourage and look forward to engaging with other teams as they continue to expand our understanding of this important topic.

On the Seshat team, we adopt a low-hanging fruit approach to the selection of hypotheses to test. The development of more direct proxies for affluence might be a good example of this since the

proposals of Baumard & Boyer regarding the origins of MSP deserve further investigation. Likewise, we appreciate that some of the broader theoretical questions about Axiality and transcendentalism are tantalizing and worthy of deeper exploration in future research. These are examples of low-hanging fruit not only because they are well-grounded in theory but because we already have so many relevant variables that it may be relatively affordable to add just a few more. This low-hanging fruit approach depends on being able to bridge the divides between humanities disciplines and big data science via extensive transdisciplinary dialogue. RBB's target article-plus-commentaries-plus-response format is well designed to assist with that process.

Author contributions

H.W., J.L., and P.T. drafted the manuscript with input from all authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Peter Turchin http://orcid.org/0000-0002-1292-8100
Harvey Whitehouse http://orcid.org/0000-0001-6935-6724
Daniel Hoyer http://orcid.org/0000-0003-2675-257X
R. Alan Covey http://orcid.org/0000-0002-1271-230X
John Baines http://orcid.org/0000-0003-4543-9909
Andrey Korotayev http://orcid.org/0000-0003-3014-2037
Andrea Squitieri http://orcid.org/0000-0002-6746-944X

References

Assmann. (2010). *The price of monotheism* (Robert Savage, Trans.). Stanford CA: Stanford University Press.
Baumard, N., & Boyer, P. (2013). Explaining moral religions. *Trends in Cognitive Sciences*, *17*(6), 272–280. https://doi.org/10.1016/j.tics.2013.04.003
Baumard, N., Hyafil, A., Morris, I., & Boyer, P. (2015). Increased affluence explains the emergence of ascetic wisdoms and moralizing religions. *Curr Biol*, *25*(1), 10–15. https://doi.org/10.1016/j.cub.2014.10.063
Beheim, B., Atkinson, Q., Bulbulia, J., Gervais, W.M., Gray, R., Henrich, J., Lang, M., Monroe, M.W., Muthukrishna, M., Norenzayan, A., Purzycki, B.G., Shariff, A., Slingerland, E., Spicer, R., & Willard, A.K. (2019). Corrected analyses show that moralizing Gods precede complex societies but serious data concerns remain. *PsyArXiv*. https://doi.org/10.31234/osf.io/jwa2n
Bellah, R. N. (2011). *Religion in human evolution: From the Paleolithic to the Axial Age*. Harvard University Press.
Boehm, C. (2008). A biocultural evolutionary exploration of supernatural sanctioning. In J. Bulbulia, R. Sosis, E. Harris, R. Genet, & K. Wyman (Eds.), *Evolution of religion: Studies, theories, and critiques* (pp. 143–152). Collins Foundation Press.
Boyer, P., & Lienard, P. (2007). Precaution systems and ritualized behavior. *Behavioral and Brain Sciences*, *29*(6), 635–650. https://doi.org/10.1017/S0140525X06009575
Geertz, A. W. (2014). Do big gods cause anything? *Religion*, *44*(4), 609–613. https://doi.org/10.1080/0048721X.2014.937052
Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*, *37*(5), 701–713. https://doi.org/10.1177/0146167211400208

Kazen, T. (2014). The role of disgust in priestly purity law. *Journal of Law, Religion and State*, 3(1), 62–92. https://doi.org/10.1163/22124810-00301004

Ketelaar, T., & Au, W. T. (2003). The effects of guilty feelings on the behavior of uncooperative individual in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion*, 17(3), 429–453. https://doi.org/10.1080/02699930143000662

Laertius, D. (1925). *Lives of Eminent Philosophers, Volume II: Books 6-10* (R. D. Hicks, Trans.). Loeb Classical Library 185. Cambridge, MA: Harvard University Press.

Lee, S. W. S., & Schwarz, N. (2011). Wiping the slate clean: Psychological consequences of physical cleansing. *Current Directions in Psychological Science*, 20(5), 307–311. https://doi.org/10.1177/0963721411422694

Lee, S. W. S., & Schwarz, N. (2021). Grounded procedures: A proximate mechanism for the psychology of cleansing and other physical actions. *Behavioral and Brain Sciences*, 44(e1), 1–69. https://doi.org/10.1017/S0140525X20000308

Levine, J., & Haar, B. T. (2019). East Asia: North China. In D. Hoyer, & J. Reddish (Eds.), *Seshat history of the axial age* (pp. 181–196). Beresta Books.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and STAN* (2nd ed.). CRC Press.

McKay, R., Herold, J., & Whitehouse, H. (2013). Catholic guilt? Recall of confession promotes prosocial behavior. *Religion, Brain, and Behavior*, 3(3), 201–209. https://doi.org/10.1080/2153599X.2012.739410

McKay, R., & Whitehouse, H. (2014). Religion and morality. *Psychological Bulletin*, 141(2), 447–473. https://doi.org/10.1037/a0038455

Mullins, D. A., Hoyer, D., Collins, C., Currie, T., Feeney, K., François, P., Savage, P. E., Whitehouse, H., & Turchin, P. (2018). A systematic assessment of 'Axial Age' proposals using global comparative historical evidence. *American Sociology Review*, 83(3), 596–626. https://doi.org/10.1177/0003122418772567

Naether, F. (2021/2020). Some remarks on Whitehouse et al. (2019), 'Complex societies precede moralizing gods throughout world history': A qualitative review of ancient world data (ancient Egypt, Roman empire) from Seshat: The global history databank. *Journal of Cognitive Historiography*, 6(1-2), 113–121. https://journal.equinoxpub.com/JCH/article/view/21159

Norenzayan, A. (2013). *Big Gods: how religion transformed cooperation and conflict*. Princeton University Press.

Nettle, D., & Frankenhuis, W. E. (2020). Life-history theory in psychology and evolutionary biology: one research programme or two? *Philosophical Transactions of the Royal Society B*, 375, 20190490. http://dx.doi.org/10.1098/rstb.2019.0490

Norenzayan, A., Shariff, A. F., Willard, A. K., Slingerland, E., Gervais, W. M., Mcnamara, R. A., & Henrich, J. (2016). The cultural evolution of prosocial religions. *Behavioral & Brain Sciences*, 39, 1–65. https://doi.org/10.1017/S0140525X14001356

Parker, R. (1983). *Miasma: Pollution and purification in early Greek religion*. Clarendon Press.

Pellò, C. (2018). The lives of Pythagoras: A proposal for reading Pythagorean metempsychosis. *Rhizomata*, 6(2), 135–156. https://doi.org/10.1515/rhiz-2018-0007

Reddish, J., & De Angelis, F. (2019). Mediterranean basin: The Greco-Roman world. In D. Hoyer, & J. Reddish (Eds.), *Seshat history of the axial age* (pp. 67–89). Beresta Books.

Reddish, J., & Manning, J. G. (2019). North Africa: Egypt. In D. Hoyer, & J. Reddish (Eds.), *The Seshat history of the axial age* (pp. 217–240). Beresta Books.

Ross, R. M., van Aert, R. C. M., van den Akker, O. R., & van Elk, M. (2021). The role of meta-analysis and preregistration in assessing the evidence for cleansing effects. *Behavioral and Brain Sciences*, 44(e19), 45–46. https://doi.org/10.1017/S0140525X20000606

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, 20(4), 523–528. https://doi.org/10.1111/j.1467-9280.2009.02326.x

Schlieter, J., & Rota, A. (2022). War, "the Father of All"—including moralizing religions? *Religion, Brain & Behavior*. https://doi.org/10.1080/2153599X.2022.2065347

Stausberg, M. (2014). Big Gods in review: Introducing Ara Norenzayan and his critics. *Religion*, 44(4), 592–608. https://doi.org/10.1080/0048721X.2014.954353

Strathern, A. (2019). *Unearthly powers: Religious and political change in world history*. Cambridge University Press.

Thom, J. C. (2020). *The Pythagorean Acusmata*. Oxford University Press.

Turchin, P., Hoyer, D., Korotayev, A., Kradin, N., Nefedov, S., Feinman, G., Levine, J., Reddish, J., Cioni, E., Thorpe, C., Bennett, J. S., Francois, P., Whitehouse, H. (2021). Rise of the war machines: Charting the evolution of military technologies from the Neolithic to the Industrial Revolution. *PLOS ONE*, 16(10), e0258161. https://doi.org/10.1371/journal.pone.0258161

Turchin, P., Whitehouse, H., Gavrilets, S., Hoyer, D., François, P., Bennett, J. S., Feeney, K., Peregrine, P., Feinman, G., Korotayev, A., Kradin, N., Nefedov, S., Levine, J., Reddish, J., Wacziarg, R., Mendelson-Gleason, G., & Benam, M. (2022). *Disentangling the evolutionary drivers of social complexity in human history: A comprehensive test of hypotheses*. https://doi.org/10.31235/osf.io/tekb6.

van Houte, M. (2022). Determinism and divine intervention: divine punishment in stoic theology. In E. Pachoumi (Ed.), *Conceptualising Divine unions in the Greek and Near Eastern worlds* (pp. 269–290). Brill.

Whitehouse, H., François, P., Cioni, E., Levine, J., Hoyer, D., Reddish, J., & Turchin, P. (2019). Conclusion: Was there ever an Axial Age? In D. Hoyer, & J. Reddish (Eds.), *The Seshat history of the Axial Age* (pp. 395–408). Beresta Books.

Whitehouse, H., François, P., Savage, P. E., Currie, T. E., Feeney, K. C., Cioni, E., Purcell, R., Ross, R. M., Larson, J., Baines, J., ter Haar, B., Covey, A., Turchin, P. (2021). Retraction note: Complex societies precede moralizing gods throughout world history. *Nature*, *595*(7866), 320. https://doi.org/10.1038/s41586-021-03656-3

Zhmud, L. (2012). *Pythagoras and the Early Pythagoreans*. Oxford University Press.

Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, *313*(5792), 1451–1452. https://doi.org/10.1126/science.1130726