nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-020-16035-9 OPEN

Scale and information-processing thresholds in Holocene social evolution

Jaeweon Shin[1], Michael Holton Price[2], David H. Wolpert[2,3✉], Hajime Shimao[2], Brendan Tracey[2] & Timothy A. Kohler[2,4,5,6✉]

Throughout the Holocene, societies developed additional layers of administration and more information-rich instruments for managing and recording transactions and events as they grew in population and territory. Yet, while such increases seem inevitable, they are not. Here we use the Seshat database to investigate the development of hundreds of polities, from multiple continents, over thousands of years. We find that sociopolitical development is dominated first by growth in polity scale, then by improvements in information processing and economic systems, and then by further increases in scale. We thus define a Scale Threshold for societies, beyond which growth in information processing becomes paramount, and an Information Threshold, which once crossed facilitates additional growth in scale. Polities diverge in socio-political features below the Information Threshold, but reconverge beyond it. We suggest an explanation for the evolutionary divergence between Old and New World polities based on phased growth in scale and information processing. We also suggest a mechanism to help explain social collapses with no evident external causes.

[1] Department of Mathematics, Rice University, 6100 Main St, Houston, TX 77005, USA. [2] Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA. [3] Center for Biosocial Complex Systems, Arizona State University, Tempe, AZ 85281, USA. [4] Department of Anthropology, Washington State University, Pullman, WA 99164-4910, USA. [5] Crow Canyon Archaeological Center, 23390 C R K, Cortez, CO 81321, USA. [6] Research Institute for Humanity and Nature, 457-4 Kamigamo Motoyama, Kita-ku, Kyoto 603-8047, Japan. ✉email: dhw@santafe.edu; tako@wsu.edu

At any single moment over the last 10,000 years, there has been great variation in how the human societies of the world are organized. Despite this cross-sectional heterogeneity, striking regularities appear to govern the evolution of societies through time. The domestication of plants and animals and formation of sedentary agricultural villages developed independently in about a dozen separate regions, in both the Old and New Worlds, following climatic stabilization at the end of the last Ice Age[1–3]. In many of these societies, domestication of plants and animals combined with changing human mobility patterns to generate rapid population growth[4,5]. Rather similar processes of urbanization eventually followed this population growth in many parts of both the Old and New Worlds[6–8]. In most societies, wealth inequality increased with greater reliance on food production, with the development of technologies yielding more surplus, with more efficient wealth transmission between generations, and with more prominent political hierarchies and various other factors promoting economic defensibility[9–11].

More quantitatively, there is pronounced covariation among the values of many variables measuring various aspects of society when assessed over long periods. Productive agricultural systems, large group sizes, urbanism, political hierarchy, and high levels of wealth inequality often—though not completely—coincide. In every region of the world permitting agriculture, we find considerable similarity in both the beginning forms of human social groups, as relatively small-scale hunter-gatherer groups, and their ultimate (or at least current) forms, as large-scale urban societies. All the interesting variability—the history—lies in the differing ways that the societies in each region develop from those beginnings.

To determine whether these shared patterns in the dynamics of societies through time reflect propinquity and borrowing, shared roots, or entrainment along common paths via competitive pressures requires detailed data on their development around the world, stretching back at least to the end of the Pleistocene. Fortunately, our knowledge of Holocene prehistory and history has expanded rapidly in the last several decades, facilitated by the requirement in many countries for site investigation prior to destruction by development[12], by increased chronological accuracy (e.g., ref. [13]), and by improvements in the scope and precision of various climatic proxies that permit better understanding of the role of climatic variability in sociopolitical change[14]. Due to the rapid accumulation of such data, once-authoritative comparative studies (e.g., refs. [15,16]) deserve continuous re-evaluation, especially since these earlier studies emphasized the more recent, historical end of the spectrum of increasing sociopolitical complexity.

One notable project addressing this demand has resulted in a dataset called "Seshat: Global History Databank"[17]. This is an ambitious coding of more than 1500 variables containing state-of-the-art knowledge on variables related to social complexity for hundreds of polities. The polities in the Seshat database for which comprehensive data are available span six continents and date from the Neolithic to the middle of the last millennium.

This dataset has enabled a number of recent publications[17–21] (see also Supplementary Note 1: Overview of Seshat and Supplementary Note 2: Previous Seshat Research). Particularly important here, Peter Turchin and colleagues[22] combined 51 variables recorded in Seshat to produce nine new variables, which they call Complexity Characteristics (CCs). They then performed a principal component analysis (PCA) of 285 polities (as of March 25, 2020, the Seshat database contains 291 polities) in that nine-dimensional CC space, as described in more detail below. They found that the first principal component (PC/PC1) captures 77% of the underlying variation in the dataset. In addition, they found that polities almost always increase their PC1 value with time. This led them to portray sociopolitical evolution as a largely homogeneous process through time in which scalar growth (in population, capital population, and territory size) and increase in the number of levels and sophistication of government, infrastructure, writing, texts, and exchange media all occur in a highly correlated fashion.

The PCA of Turchin and colleagues answers the question, "what is the largest dimension of shared variability across all the variables and among all the cases in the Seshat data?" where these cases range from village-level societies to empires. It is important to note that this question is essentially ahistorical, not only disembedding the cases (a particular polity in a particular century) from their regions, but also from their own trajectories of development, and from that of similar polities. In contrast, here we are more interested in questions like "how do polities become more complex as they evolve?" and "what paths do polities follow through the Seshat feature space?"

However, the fact that the first principal component of a PCA by itself captures a large amount of the large-scale variation in the data does not preclude there being rich structure relating the components of the (vector-valued) data at a more fine-grained scale. To give a stark example, suppose that we have a two-dimensional dataset, generated by sampling the two equations $x(t) = t$, $y(t) = \sin(t)$. (So t is a hidden variable, not recorded in the data.) If the data come from a long-enough range of t values, then a PCA will tell us that PC1 is (approximately) identical to the x direction, while PC2 is identical to the y direction, and that PC1 "captures a large amount of the variation in the data". Yet as shown in Fig. 1, in fact there is far more to understand about the relationship between PC1 and PC2, namely that $y = \sin(x)$ exactly. Ironically, the fundamental issue we raise is very closely related to one identified by Turchin and Korotayev who noted that correlational analyses of cross-cultural data yield misleading results when two variables (population density and incidence of warfare, in their case) are linked in a dynamical system and oscillate relative to each other with time lags, as in a predator-prey system[23]. In the case of the correlational analysis in Turchin et al.[22], the issue is not with time lags, though, but with changes in

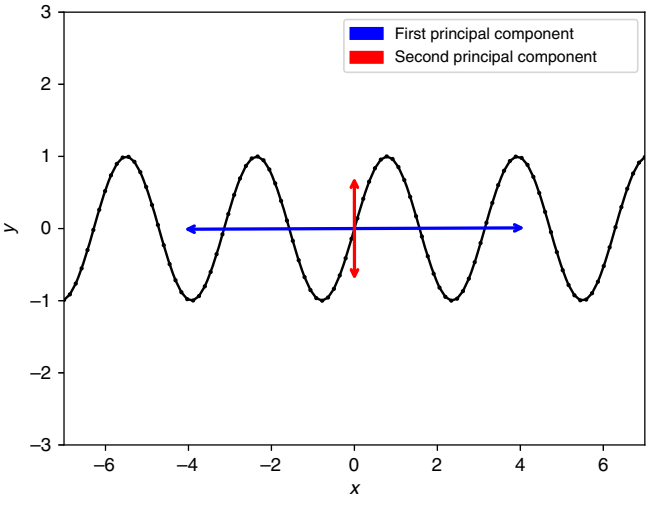


Fig. 1 PCA decomposition of the data generated from sinusoidal function. Datapoints on the plot are drawn at uniform interval from parametric equations $x(t) = t$ and $y(t) = \sin(2t)$, where t ranges from −7 to 7. The blue bidirectional line represents the first Principal Component, which explains 98% of the variance in the data. The red bidirectional line represents the second Principal Component, which explains the rest of the variance in the data.

the direction of the relationship between PC1 and PC2 across the developmental ranges of the societies they describe.

Motivated by such considerations, in this article we analyze the covariation between the values of the top two principal components in the Seshat dataset. We find that there is a striking, highly patterned relationship between them—similar to the sine wave relationship of Fig. 1, in fact.

To help understand the social and historical significance of this relationship, we examine how the weightings (loadings) of the CCs under the vector PC1 and under the vector PC2 change as a polity develops. This provides our first result: According to the Seshat database, the development of polities on average is dominated first by a period of growth in scale, e.g., in the capital's population and territory size. After that the dynamics enters a period when it is dominated by improvements in information-processing and economic systems (i.e., by transactional and information-storage capabilities). These improvements are then followed by further increases in scale in a third period. This result suggests that to make major improvements in information processing, a polity must first surpass a scalar threshold. It also implies that the further increases in scale in the third region depend in turn on improvements in information-processing ability. Social evolution is thus a contingent, historical process, but one with strong constraints on its form imposed by structural considerations (in the sense of ref. [24]).

We interpret the two transitions between these three periods as permeable boundaries in social complexity. The first boundary moving from left to right along increasing values for PC1, which we call the Scale Threshold, separates those polities that have undergone more growth in scale than in information-processing capacity, from those polities that have already gone through such scale growth and are now differentially increasing their information-processing capacity. The second boundary, which we call the Information Threshold, separates polities that have not achieved the information-processing capacity that appears to be needed for additional growth in scale, and so are still growing that capacity, from those polities that have in fact achieved sufficient information-processing capacity for further increases in scale.

Next we investigate the dynamics of the NGAs recorded in the Seshat dataset, making full use of the century-resolution time-stamps provided in Seshat. This provides our second, albeit qualitative result: There appears to be a first region in PC1–PC2 space in which polities are relatively concentrated. This is followed by a second region where polities become more spread apart; we can speak of them taking different paths with respect to information processing as they grow in scale. In a final, third region, there appears to be a pronounced homogenization of features in which polities come to lie almost on top of one another. We note that this final convergence might simply reflect saturation of the values of the particular features that are recorded in Seshat, which were chosen with pre-modern polities in mind. Intriguingly, the first region roughly coincides with the scale-dominated growth described above, before the Scale Threshold is crossed. The second region coincides with a more variable dynamics, where many but not all polities achieve improvements in information-processing (in the sense discussed above). The third region roughly coincides with renewed emphasis on growth in scale, after the Information Threshold has been crossed.

While performing these analyses we uncovered patterns that turned out to be statistical artifacts. In particular, there is a pronounced clustering of the PC1 values in two distinct regions, which can be accurately described as a mixture of two Gaussians of very similar widths (see Supplementary Note 3: Bimodality and Supplementary Figs. 1–4). Such clusters in a sociopolitical feature space arise often in sets of data on long-term cultural evolution, and are sometimes interpreted as basins of attraction of an underlying dynamics. However, as we note in the Supplementary Note 4 (Possible non-social-science explanations of the bimodality; see also Supplementary Figs. 5 and 6), the clustering in Seshat appears to be an artifact of a type not so far described in the literature, to our knowledge.

This provides a cautionary tale on analyzing datasets that combine samples of multiple, finite duration runs of the same Markov chain, if (as in Seshat) each run has a different starting position and a different starting time.

We now turn to presenting our results on the interaction between PC1 and PC2, with particular emphasis on how societies (henceforth, polities) move through the space jointly defined by these dimensions. We then discuss some implications of our findings for the role of population increase in social evolution, the divergence of polities in the Old and New Worlds, and possible new pathways for explaining social instability these results suggest.

Results and discussion

Interaction between PC1 and PC2. We find a pronounced structure relating the values of PC1 with the values of the other eight principal components, despite PC1 containing 77% of the covariation across polity feature vectors. Here we report the relationship between the first and second principal components, PC1 and PC2. Figure 2 shows the average of polity values in PC2 evaluated in a sliding window along PC1 space. It is clear that a non-linear interaction between PC1 and PC2 exists, where the average PC2 scores of polities first drop, then rise, and then drop again as one moves along PC1. In other words, there is an interdependent relation between PC1 and PC2 despite no Pearson correlation in aggregate.

The first hinge point in Fig. 2 occurs near the PC1 value of −2.5 where average PC2 values abruptly stop decreasing and start to increase. Polities located near this point include Upper Egypt ca. 3600 BC, the Paris Basin ca. 400 BC (La Tène A), the Iceland Commonwealth ca. AD 1000, and the Valley of Oaxaca at the same time.

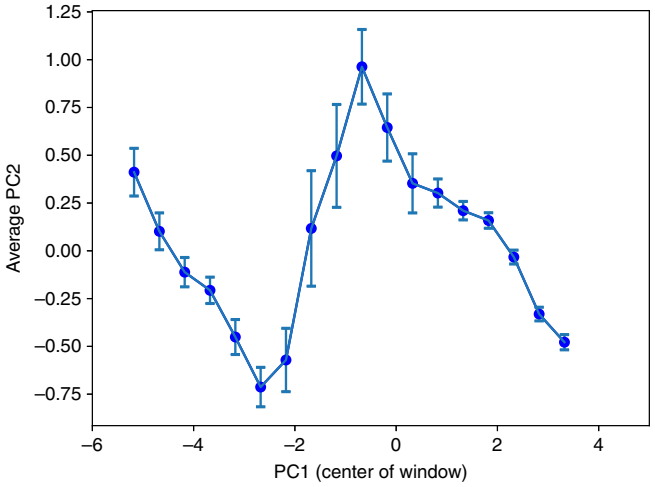


Fig. 2 The average score of observations on PC2 in a sliding window along PC1. The PCA is based on 414 datapoints. Each PC1 sliding window is defined to have width 1.0 (on the scale of PC1) and overlap width of 0.5 with other sliding windows, i.e., the center of the PC1 sliding window occurs at every 0.5 interval. The error bars are defined as the mean(PC2) ± SE (PC2), i.e., mean of the PC2 values within the sliding PC1 windows +/− standard error of the PC2 values. Number of samples in each sliding window, left to right, is 12, 25, 48, 50, 37, 35, 28, 20, 21, 22, 33, 35, 51, 89, 104, 86, 64, 42.

Table 1 Loadings of the nine complexity components onto PC1 and PC2, and the percentage of variance of the full dataset explained by each PC. We present our calculations which differ very slightly from those presented in ref. [22].

	PolPop	PolTerr	CapPop	levels	gvrnmt	infrastr	writing	texts	money	%Var
PC1	0.93	0.85	0.89	0.90	0.88	0.88	0.87	0.92	0.80	77.2
PC2	−0.25	−0.35	−0.27	−0.15	0.09	0.12	0.31	0.24	0.28	6.0

The second hinge point in Fig. 2 occurs at a value for PC1 of around −0.5, where average values of PC2 suddenly switch from increasing to decreasing as PC1 increases. Polities located near this hinge include proto-Elamite Susiana ca. 3000 BC, and First Intermediate period Upper Egypt ca. 2100 BC.

It is important to be clear about how the changes in direction in Fig. 2 should be interpreted. Movement to the right along the PC1 axis happens as polities increase their values of those CCs that have positive loadings on PC1 while decreasing those that have negative loadings. Due to the choice of CCs in ref. [22] to be quantities that one would expect to increase together, they all have positive loadings on PC1 (though led most strongly by the CCs that the Seshat database labels *PolPop*, *texts*, *levels*, and *CapPop*—see top row in Table 1). Since PC2 must be orthogonal to PC1, this means that PC2 must both have CC components with positive loadings and those with negative loadings (see second row in Table 1). The movement downwards in PC2 on the left-hand side of Fig. 2 reflects increasing values on those CCs that have negative loadings in PC2—led by *PolTerr*, *CapPop*, and *PolPop*—while values on CCs with positive loadings in PC2 (led by *writing*, *money*, and *texts*) increase more slowly or not at all. On the other hand, movement upwards along PC2 happens when values for those CCs with positive loadings on PC2 are differentially increasing more than the others. Together these three CCs of *money*, *writing*, and *texts* with positive loadings (and less importantly those with lower positive scores on PC2 in Table 1) drive the changes in direction along the PC2 axis in Fig. 2. They reflect what we call the information-processing capacity of a polity.

These changes in which CCs drive the dynamics along PC2 are illustrated in Fig. 3, which decomposes PC2 into two sets of CCs, ones with positive loadings on PC2 (*money*, *writing*, *texts*, *infrastructure*, and *gvrnmt*) and others with negative loadings (*PolTerr*, *CapPop*, *PolPop*, and *levels*). Again, the CCs with positive PC2 loadings are related to information-processing capabilities, while those with negative loadings are related to scale. For each PC1 value, the red and blue lines represent the contribution of the two sets of CCs to PC2 (the sum of the product of the CC values and their loading on PC2). In other words, the red (blue) line indicates the PC2 value evaluated only using the information- (scale-) related CCs. The black line is the summation of the red and blue, and is equivalent to the line in Fig. 2.

As polities grow in their PC1 value, the relative contributions of the scale CCs and the information-processing CCs changes. Up to the first hinge point at PC1 ~−2.5, polities grow substantially in scale as their PC1 value increases (i.e., the blue line drops), with relatively less growth in their information-processing capabilities. Since the PC2 vector has negative components for those scale-related CCs, this results in declining values on PC2. Between −2.5 and −0.5 on PC1, movement along PC1 is accompanied by relatively rapid increases in information-processing capabilities (though the error bars in Fig. 2 indicate more heterogeneity in scores on PC2 in this portion of PC1 than in any other portion). Since the PC2 vector has positive components for the information-processing CCs, this results in an increasing average score for polities on PC2 as PC1 grows in this region. Then as

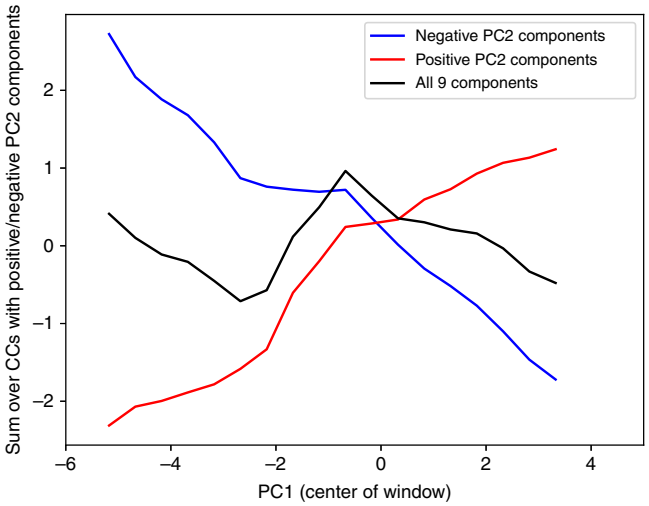


Fig. 3 The decomposition of the average PC2 into two subsets of CCs. The red line represents the CC components related to information-processing capabilities (*writing*, *money*, *texts*, *infrastructure*, and *gvrnmt* in order of importance), having positive correlations with PC2. The blue line represents the CC components primarily related to scale (*PolPop*, *levels*, *CapPop*, and *PolTerr* in order of importance). Note the changes in gradients along these two lines as they traverse PC1. The black line here is identical to the blue line in Fig. 2, with a change of scale.

polities move past −0.5 in PC1, again growth in scale CCs dominates the change in information-processing CCs, resulting in the decreasing values of PC2 in Fig. 2.

These dynamics suggest that at a level of complexity represented by values of PC1 of about −2.5, growth in scale may become constrained in polities where writing, money, or texts are relatively unsophisticated (store little information). Developing increased capacities in these areas appears to allow further increases in scale. This is why we suggest that that hinge point can be interpreted as a Scale Threshold marking an important transition in the dynamics of developing polities. At PC1 values of about −0.5 another, Information Threshold, is passed, appearing to facilitate further increases in scale.

Dynamics in the PC1–PC2 space. It is possible that the results in Figs. 2 and 3 could be driven by a few aberrant cases, perhaps a few large societies of the Americas that lack writing or some small societies that have writing for some reason like borrowing, or because they have budded off from a larger society. (In fact, all these possibilities were raised by a reviewer.) To help address such issues, in Fig. 4 we present a more fine-grained picture of the Seshat PC1–PC2 values, by including the time-stamps of the datapoints to depict trajectories of individual NGAs. Arrows run from an entry in Seshat giving the (PC1, PC2) coordinates of a particular polity at a particular time to the (PC1, PC2) coordinates of the next recorded time for that same polity.

Typically, the time between such datapoints is 100 years, but often there are gaps in the sequence[22]. For observations more than 100 years apart, the dots in Fig. 4 show the interpolated

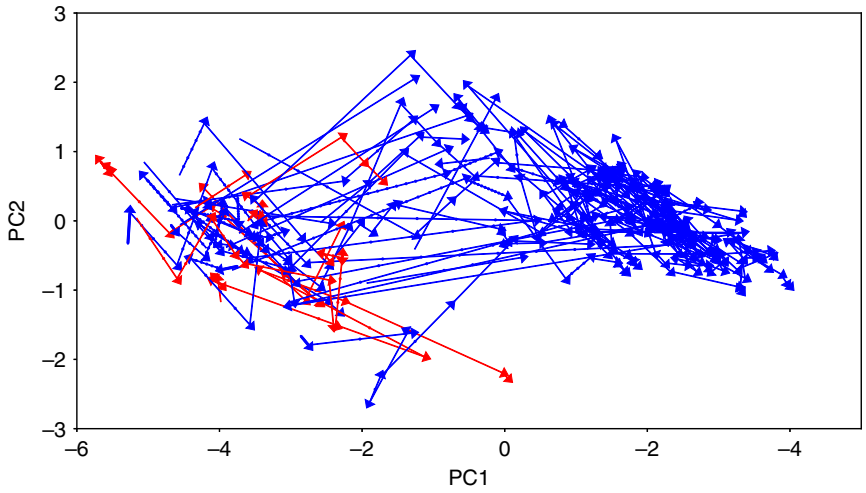


Fig. 4 The trajectories of individual NGAs in PC1–PC2 space. Red indicates New World NGAs; blue indicates Old World NGAs.

locations along each arrow at a spacing of 100 years, given a linear interpolation between the points $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$. Red arrows stand for New World polities, and blue for those in the Old World.

Figure 4 suggests several features of sociopolitical evolution that have been previously poorly recognized (if at all). First, as mentioned above, the thresholds picked out by the hinge points in Fig. 2 approximately coincide with other changes in the dynamics through PC1–PC2 space. For example, societies lying between the Scale and the Information thresholds are more heterogeneous in their movement in both PCs than are societies lying to the right of the Information Threshold. Movement past that threshold, into the right-most region, seems to be in part a process of homogenization of features.

Second, the movement upward in the PC2 dimension between values of about −2.5 to near 0 on PC1 is nearly universal in the Old World regions analyzed in ref. [22]—though not in the Americas. (The six NGAs analyzed in the New World are the Finger Lakes, Cahokia, and Valley of Oaxaca in North America, and the Lowland Andes, North Colombia, and Cuzco in South America. The complete sample is listed in Turchin et al. [ref. [22], Table SI1]; see also Supplementary Fig. 10). Evidently, New World behavior is indeed unusual—but works to weaken the patterns in Figs. 2 and 3 rather than strengthen them. We leave for the future more work on the relationship between New World polities and the Scale Threshold, especially since the sample from Seshat analyzed here omits key New World societies, such as the Maya, Teotihuacan and the Aztec.

In the current sample, New World societies are also unusual in that they are largely confined to a loose left-hand cluster, centered around (P1 = −3.5, PC2 = −0.5). Cahokia, for example, achieves its highest score on PC1 at A.D. 1200 (−2.57; PC2 = −1.21), approaching but not quite crossing the Scale Threshold. In this sample, only the Cuzco NGA under the Inca crosses the Information Threshold, during the AD 1400s and 1500s, with a PC1 score of 0.02 in A.D. 1500, but markedly low scores on PC2 (−2.26 in A.D. 1500). For the moment we conclude that societies in the New World sample may be constrained in size by their low information-processing capacities.

Perhaps the most remarkable feature in Fig. 4 though is the nearly universal movement upwards in the PC2 dimension in the middle of the PC1 range, in between the two thresholds (especially if one restricts attention to Old World societies). This movement is what causes the change in direction of the line in Fig. 2 (To reiterate the point made above, this reflects differential improvements in information processing in the middle zone of PC1 than in either of the other two zones.).

The general downwards slope of most trajectories for positive values of PC1, as polities pass the Information Threshold, is also noteworthy. In the part of the space with even greater PC1 values it is tempting to conclude from informal examination of the plot, and its dense cluster of points around (PC1 = 2, PC2 = 0), that there is a sink in the middle of the right cluster. While possible, such a cluster would need to be confirmed with further statistical analysis.

A final point of interest demonstrated in Fig. 4 is that human societies do not exist in most of the possible space of joint social and cultural traits. The later cluster is even more focused on a portion of the available space than are the societies in the looser, earlier cluster. We do not (for example) see societies with sophisticated information processing but small size. This suggests that it is impossible for human societies to arise (let alone persist) unless they satisfy a tight set of constraints limiting their sociopolitical characteristics.

Thesholds and morphospace diversity. To a first approximation, the Seshat database analyzed here and in ref. [22] demonstrates a shared movement through time by all the polities in this sample along an axis (PC1) representing simultaneous increases in scale and in a number of indicators of governmental, infrastructural, informational, and economic capacities. This homogeneous process is characterized as having produced a "striking similarity in the way that the societies in our global historical sample are organized" ([ref. [20], p. 4 of 8]; see also [ref. [22], p. 3]).

Going beyond this first approximation, however, we demonstrate a pronounced structure relating the scale and information complexity characteristics recorded in Seshat. This structure strongly suggests that sociopolitical evolution is initiated primarily by growth in scale. Only once a certain size is reached, which we call the Scale Threshold, does growth in information-processing capacity start to accelerate relative to growth in scale. This suggests that a polity's growth in scale may become limited by low capacity in its writing, exchange and similar information-processing systems. Once those capacities reach a certain level, polities can cross what we call the Information Threshold, allowing further growth in scale. We emphasize of course that these thresholds are not Iron Curtains separating incommensurable worlds, but elastic and porous frontiers between zones in which different constraints are most pronounced.

Moreover, the leftmost zone in Figs. 2–4, corresponding to the first stage of this dynamics, seems to be one in which different polities develop in idiosyncratic ways, diverging in the space of sociopolitical features. The second stage, dominated by growth in information-processing capabilities, seems to be one in which polities—with the notable exception of those in the New World—all start moving in parallel. Then in the third stage—across the Information Threshold—polities converge in feature space. We caution that these patterns are all based on the current Seshat database which, impressive as it is, is still quite small.

Peregrine[25] reports an analysis of the archaeological traditions coded in ref. [26] in which factor analysis was used to define a two-dimensional morphospace within which each of the traditions could be located at a specific time. As in our analysis of Seshat, a great deal of the available morphospace was never (or rarely) occupied. Following ref. [27], Peregrine attributes some of this empty space to functional constraints (for example, were small-scale societies to undertake construction of large-scale public monuments, it could only be to the [lethal] detriment of their day-to-day food-getting activities). Other empty space is likely due to developmental constraints (for example, small-scale societies may simply not have enough interacting minds to innovate and maintain complex technologies).

The observation that societies focus on a smaller portion of the possible space defined by PCs 1 and 2 as they move into the right-most cluster is reminiscent of the common observation that small-scale societies are relatively more diverse than large-scale societies in terms of both cultural practices and organization (e.g., [ref. [28], pp. 8–10]). Note in particular that in the part of the transition region where scores for PC1 hover around 0, polities with low scores on PC2 tend to move up, while polities with high values on the second PC tend to move down (Fig. 4). A plausible mechanism for this convergence that deserves further examination is that the selective environments inhabited by societies become increasingly competitive as they move to the right along the PC1 dimension. Highly selective environments reward societies with constellations of traits that allow them to survive and spread. Diverse societies either develop similar constellations of traits through this selection process, or they fail to survive. Similar dynamics are predicted in refs. [29,30].

Three points raised by our analysis warrant further discussion. The first has to do with how social evolution has typically been explained. The second concerns the Old/New World divergence revealed here. Third, this analysis points to new avenues for understanding social instability. We address these in turn below.

Role of population. Archaeologists have long held different views on the importance of population growth in social evolution. From a perspective grounded in the small- and middle-range societies in the US Southwest, for example, Kohler and Crabtree[31] have recently argued that increasing social and political complexity is largely driven by population growth. From a point of view more connected with the ancient states of Mesoamerica, however, Feinman[32] recognizes relatively strong correlations between population size and sociopolitical complexity, but terms the relationship "messy," arguing that it is significantly mediated by the "specific nature of the relational links between individuals" especially the specific organization of the institutions by which different societies are governed [ref. [32], p. 47]. Our findings here—that growth in scale is dominant at the extrema of PC1 but that development of information processing and exchange mechanisms is critical in the middle-range of PC1—suggest that both of these perspectives can be accommodated.

Our results here are in general accord with the findings of an earlier body of research in anthropology on scalar stress that modeled a relationship between increasing group size and the appearance of hierarchical structures facilitating information processing[33,34]. Such research provides the uniformitarian microfoundations for our own work, via studies of small groups, variability in spans of control in organizations of various types, and ethnographies. We focus instead on understanding how this relationship played out in the global history of Holocene societies.

Moreover, our analysis raises the possibility that the zone between the thresholds of Scale and Information is especially critical for the evolution and elaboration of institutions. Institutions are mechanisms "whose outcomes are rules of interactions" and they "provide a means for individuals to amalgamate dispersed information about resources and wants, and hence coordinate … actions to reach an equilibrium that gives higher pay-offs" than would be possible in their absence [ref. [35], p. 3–4]. Institutions maintain and promulgate information about peoples' past behaviors—information that becomes more difficult to recover as group sizes grow [ref. [35], p. 4]. So institutions thrive on and deal in information, and the actors controlling them will seek to improve the mechanisms for its storage and retrieval.

For example, Koji Mizoguchi describes Japan's Kansai region ca. A.D. 400 (Middle Kofun, PC1 score −0.57) as experiencing an increasing number of distinct fields of communication resulting in part from extension of political interests into neighboring regions of Korea, coupled with attempts to institutionalize communications. By A.D. 500 (late Kofun, PC1 score 0.20) this region has crossed the Information Threshold into an "archaeology of bureaucracy" in which the elite were paving the way for "registration and direct control of commoners" [ref. [36], p. 308]. Within a century, by A.D. 600 (Asuka, PC1 score 1.64) the Kansai region hosted the rise of the palace, the introduction of Buddhism, the first national treasury, the introduction of Chinese-modeled fiscal policies and script, and formalization of the *Ritsuryō* law code. The fifth and sixth centuries likewise saw significant spatial expansion in the area controlled by the dominant clan, presumably enabled, in part, by these advances.

Diverging old and New World trajectories. One unexpected finding is that few New World societies seem to cross the Scale Threshold. We infer that as a result they do not undergo as much pressure to develop information and exchange systems as societies that did. This suggests that if we want to understand the rarity of writing systems in the New World and the absence of coinage we first need to understand the factors limiting scale in many such societies. There are many candidates, including perhaps the absence in the New World of an inland sea providing an efficient means of linking societies on its periphery.

More promising as an explanatory factor though in our view is the general absence of animals capable of carrying people or sizeable loads in the Americas. This is also of course connected with the famous absence of the wheel as a practical device. Such animals, of course, dramatically reduce the friction of transport and facilitate expansion of empire. It is notable that the only partial exception in the Seshat sample is the Cuzco NGA where camelids could carry loads, but not warriors. Not coincidentally we think, this is also the only New World society in this sample that crosses the Scale Threshold, and perhaps as a result the only to develop a recording system (the *quipu*) beyond mnemonic devices.

In view of these results, the question "Why didn't indigenous North American societies develop writing systems?" seems ill-posed. The appropriate question is perhaps "Why didn't these societies develop in scale to the degree where writing systems would have been advantageous or required?"

Explaining social instability. In most of the polities recorded in Seshat, there is a strong relationship between movement along the PC1 dimension (growth in scale coupled to growth of information-processing capacities)—but not in all of them. Could some of the frequent collapses seen in societies be due to a polity's never developing sufficient information-processing capacities, so that it stumbles or even collapses through poor performance due to lack of external connectivity, internal coherence, or inability to compete with polities whose superior information-processing abilities have enabled more growth in size? While Seshat is currently too small to address whether such "correlational failures" should be taken seriously, that may not be the case in the future.

Archaeologists have long noted the instability of societies of intermediate complexity. The Mississippian (late-prehistoric, maize-based) societies of the US Midwest and Southeast, traditionally characterized as simple or complex chiefdoms, commonly cycled unstably between these two organizational forms[37] or, in a related vision, oscillated between more dispersed and more concentrated power distributions in a near-constant churn of fission-fusion[38]. Proximate mechanisms driving this instability appear to have included both endogenous and exogenous factors: contention over chiefly succession, the fortunes of war, and climatically induced variability in the maize that fueled their tribute-based hierarchies. While such factors may in fact be sufficient, the results here suggest the further possibility that mismatch among aspects of social development should be considered as a generic cause of instability.

For example, Fried [ref. [39], p. 225] surmised that stratified societies lacking state institutions must be "one of the least stable models of organization that has ever existed" since he thought that stratified communities would have to quickly develop ever more powerful institutions of political control to maintain the differential access to basic resources that defines systems of stratification. Analyzing such problems as failures of correlation among the sociopolitical features of a polity might provide a more general view on the nature of collapses without clear exogenous triggers. Pursuing such a vision is a topic for future research.

Methods

Sample. We analyze a worldwide sample of 285 polities in 30 regions described on 51 variables relevant to sociopolitical organization by reference to archaeological and historical data, using the codings by century derived from and made available by Seshat: The Global History Project[17] (Supplementary Note 1: Overview of Seshat). Codings for each region begin as early in the Neolithic as permitted by the local archaeological record and extend through time up to just before the local industrial revolution. These 51 variables were collapsed into nine Complexity Characteristics (CCs) named in Table 1 by Turchin et al.[22] whose usage we adopt. Four of these measure aspects of social scale: Polity Population, Polity Territory, Capitol Population, and Hierarchy (i.e., number of levels in the political, military and religious establishments, and in the settlement hierarchy). The remainder either measure how information and financial transactions were processed (writing, money) or mix aspects of scale and information. Multiple imputation was used to deal with missing data, uncertainty, and expert disagreement. Turchin et al.[22] demonstrated that a first component (PC1) from a principal component analysis (PCA) explains ~77% of the variability in this dataset whereas PC2 explains ~6%; the remaining components drop rapidly in explanatory value. Turchin et al.[22] restrict their discussion of the PCA to the first principal component. Our analyses here do not overlap with more recent analyses by the Seshat team (Supplementary Note 2: Previous Seshat Research).

Analysis. We reanalyze the dataset reported by ref. [22], accepting their PCA but extending our analysis to the first two PCs. Two facts motivate and justify this extension. First, their claims about the proportions of variance explained by each PC are based on the static pattern in the data and do not take into account the dynamic development of this pattern through time, which involves moving through a non-linear relationship between PCs 1 and 2 (Fig. 2). Second, the scores on PC1 in their analysis are bimodal, whereas strict interpretation of percentages explained by each component requires unimodality[40].

We investigate this bimodality extensively in Supplementary Note 3: Bimodality. We use Gaussian mixture models to show that there are two discrete clusters in the original 9-dimensional data, and we demonstrate via bootstrapping that these are not caused by noise. We develop a novel model explained in detail in Supplementary Note 4.1: Discrete Markov Transition Model that on the whole suggests that the mode consisting of low values on PC1 is likely to be due to non-uniformity in initial values on CCs (and hence on PC1) when polities first begin to be measured by Seshat researchers. We develop some simple simulations to help judge how likely it is that the right-hand mode in the distributions of polity scores on PC1 is due to saturation in values of the variables underlying the PC. These simulations show that if saturation were responsible for the right-hand peak, this peak should occur at the positive limit in Supplementary Fig. 1, instead of at lower values of PC1.

Another way in which bimodality in the distribution of polity PC1 values could arise would be if polities between the two modes were undersampled (i.e., via selection bias). We investigate this in the Supplementary Note 4.3: Undersampling by (1) interpolating missing time periods within each region; and (2) weighting within each region depending on the sparsity of data in the neighborhood of each data point. Since bimodality survives both tests, we tentatively conclude that it is not due to selection bias.

Our main results are derived from visual analysis and interpretation of Figs. 2 and 3, and especially the dynamics displayed in Fig. 4.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available through https://github.com/jaewshin/Holocene. We used data from the Seshat databank (seshatdatabank.info) under Creative Commons Attribution Non-Commercial (CC By-NC SA) licensing (https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode). The data were accessed and downloaded in August 2018 and are identical to the dataset published and used by Turchin et al.[22]. The data contain 285 unique polities and have a total of 414 datapoints. An updated version of the dataset, which contains 291 unique polities and 864 datapoints, is available for download through http://seshatdatabank.info/datasets/.

Code availability

The code for all figures and analyses is available through https://github.com/jaewshin/Holocene.

Received: 12 November 2019; Accepted: 6 April 2020;
Published online: 14 May 2020

References

1. Richerson, P. J., Boyd, R. & Bettinger, R. L. Was agriculture impossible during the Pleistocene but mandatory during the Holocene? A climate change hypothesis. *Am. Antiquity* **66**, 387–411 (2001).
2. Bettinger, R. L. In *Macroevolution in Human Prehistory* (eds. Prentiss, A., Kuijt, I. & Chatters, J. C.) 275–295 (Springer New York, New York, 2009).
3. Kavanagh, P. H. et al. Hindcasting global population densities reveals forces enabling the origin of agriculture. *Nat. Hum. Behav.* **2**, 478–484 (2018).
4. Bocquet-Appel, J.-P. When the world's population took off: the springboard of the Neolithic Demographic Transition. *Science* **333**, 560–561 (2011).
5. Kohler, T. A. & Reese, K. M. Long and spatially variable Neolithic Demographic Transition in the North American Southwest. *Proc. Natl Acad. Sci. USA* **111**, 10101–10106 (2014).
6. Childe, V. G. The Urban revolution. *Town Plan. Rev.* **21**, 3–17 (1950).
7. Smith, M. E. V. Gordon Childe and the Urban Revolution: a historical perspective on a revolution in urban studies. *Town Plan. Rev.* **80**, 3–29 (2009).
8. Morris, I. *Why the West Rules—For Now: The Patterns of History, and What They Reveal about the Future.* (Profile Books, London, 2010).
9. Kohler, T. A. et al. Greater post-Neolithic wealth disparities in Eurasia than in North America and Mesoamerica. *Nature* **551**, 619–622 (2017).
10. Borgerhoff Mulder, M. et al. Intergenerational wealth transmission and the dynamics of inequality in small-scale societies. *Science* **326**, 682–688 (2009).
11. Mattison, S., Smith, E. A., Shenk, M. K. & Cochrane, E. E. The evolution of inequality. *Evol. Anthropol.* **25**, 184–199 (2016).
12. Sebastian, L. & Lipe, W. D. In *Archaeology & Cultural Resource Management* (eds. Sebastian, L. & Lipe, W. D.) 283–297 (School for Advanced Research Press, Santa Fe, 2009).
13. Denaire, A. et al. The cultural project: Formal chronological modelling of the Early and Middle Neolithic sequence in Lower Alsace. *J. Archaeol. Method Theory* **24**, 1072–1149 (2017).
14. d'Alpoim Guedes, J. A., Crabtree, S. A., Bocinsky, R. K. & Kohler, T. A. Twenty-first century approaches to ancient problems: Climate and society. *Proc. Natl Acad. Sci. USA* **113**, 14483–14491 (2016).

15. Adams, R. M. *The Evolution of Urban Society: Early Mesopotamia and Prehispanic Mexico* (Aldine, Chicago, 1966).
16. Trigger, B. G. *Understanding early civilizations: A comparative study* (Cambridge University Press, Cambridge, 2003).
17. François, P. et al. A macroscope for global history: Seshat global history databank, A methodological overview. *Digital Hum. Quart.* http://www.digitalhumanities.org/dhq/vol/10/4/000272/000272.html (2016).
18. Turchin, P. et al. Seshat: the global history databank. *Cliodynamics* **6**, 77–107 (2015).
19. Turchin, P. Fitting dynamic regression models to Seshat data. *Cliodynamics* **9**, 25–58 (2018).
20. Turchin, P. et al. Evolutionary pathways to statehood: old theories and new data. *SocArXiv Prepr.* https://doi.org/10.31235/osf.io/h7tr6 (2018).
21. Whitehouse, H. et al. Complex societies precede moralizing gods throughout world history. *Nature* https://doi.org/10.1038/s41586-019-1043-4 (2019).
22. Turchin, P. et al. Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proc. Natl Acad. Sci. USA* **115**, E144–E151 (2018).
23. Turchin, P. & Korotayev, A. V. Population dynamics and internal warfare: a reconsideration. *Soc. Evol. Hist.* **5**, 112–147 (2006).
24. McAllister, J. W. in *The Future of the Sciences and Humanities* (eds. Tindemans, P., Verrijn-Stuart, A. & Visser, R.) 19–54 (Amsterdam University Press, Amsterdam, 2002).
25. Peregrine, P. N. in *The Emergence of Premodern States: New Perspectives on the Development of Complex Societies* (eds. Sabloff, J. A. & Sabloff, P. L. W.) 271–295 (SFI Press, Santa Fe, 2018).
26. Peregrine, P. N. *Outline of Archaeological Traditions*. (HRAF Press, New Haven, 2001).
27. McGhee, G. R. *The Geometry of Evolution: Adaptive Landscapes and Theoretical Morphospaces*. (Cambridge University Press, Cambridge, 2006).
28. Diamond, J. *The World Until Yesterday: What can we learn from traditional societies*? (Penguin, New York, 2012).
29. Richerson, P. & Boyd, R. in *Ethnic Conflict and Indoctrination: Altruism and Identity in Evolutionary Perspectives* (eds. Eibl-Eibesfeldt, I. & Salter, F.) 71–95 (Berghahn, Oxford, 1998).
30. Turchin, P., Currie, T. E., Turner, E. A. L. & Gavrilets, S. War, space, and the evolution of Old World complex societies. *Proc. Natl Acad. Sci. USA* **110**, 16384–16389 (2013).
31. Kohler, T. A. & Crabtree, S. A. in *Peupler la Terre : De la Préhistoire à l'ère des métropoles* (ed. Sanders, L.) 173–192 (Presses Universitaires François Rabelais, Paris, 2017).
32. Feinman, G. M. in *Cooperation and Collective Action: Archaeological Perspectives* (ed. Carballo, D. M.) 35–56 (University Press of Colorado, Boulder, 2013).
33. Johnson, G. A. in *Theory and Explanation in Archaeology: the Southampton Conference* (eds. Renfrew, C., Rowlands, M. J. & Segraves, B. A.) 389–421 (Academic Press, 1982).
34. Kosse, K. Some regularities in human group formation and the evolution of societal complexity. *Complexity* **6**, 60–64 (2000).
35. Powers, S., van Schaik, C. & Lehmann, L. How institutions shaped the last major evolutionary transition to large-scale human societies. *Philos. Trans. R. Soc. B* **371**, 20150098 (2016).
36. Mizoguchi, K. *The Archaeology of Japan: From the Earliest Rice Farming Villages to the Rise of the State*. (Cambridge University Press, New York, 2013).
37. Anderson, D. G. in *Political Structure and Change in the Prehistoric Southeastern United States*, The Ripley P. Bullen Series, Florida Museum of Natural History (ed. Scarry, J.) 231–252 (University Press of Florida, Gainesville, 1996).
38. Blitz, J. H. Mississippian chiefdoms and the fission-fusion process. *Am. Antiquity* **64**, 577–592 (1999).
39. Fried, M. H. *The Evolution of Political Society: An Essay in Political Anthropology*. (Random House, New York, 1967).
40. Jolliffe, I. *Principal Component Analysis*, 2nd edn. (Springer, 2011).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. SMA-1620462 to D.H.W and T.A.K. We thank Darcy Bird, Laura Ellyson, Peter Turchin and the Seshat project for providing the dataset used here, Henry Wright, and the Santa Fe Institute and the Research Institute for Humanity and Nature for support.

Author contributions

D.H.W, H.S., M.A.P., and T.A.K. designed the research. T.A.K., D.H.W., and M.H.P. wrote the main text. J.S., H.S., B.T., M.H.P., D.H.W., and T.A.K. wrote the SI. J.S., H.S., M.H.P., and B.T. conducted the analyses and prepared the figures.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16035-9.

Correspondence and requests for materials should be addressed to D.H.W. or T.A.K.

Peer review information *Nature Communications* thanks Peter Peregrine and Simon Powers for their contribution to the peer review of this work. Peer reviewer reports are available.

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

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

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

© The Author(s) 2020

Supplementary Information for "Scale and Information-Processing Thresholds in Holocene Social Evolution" by Jaeweon Shin et al.

Contents

1 Supplementary Note 1: Overview of Seshat **1**
1.1 Choice of NGAs . . . 2
1.2 Choice of Variables and Byproducts of Imputation . . . 2
2 Supplementary Note 2: Previous Seshat Research **3**
2.1 Evolution of Information Systems . . . 3
2.2 Evolution of Specialized Governance Structures . . . 4
3 Supplementary Note 3: Bimodality **4**
3.1 Gaussian Mixture Model . . . 5
4 Supplementary Note 4: Possible non-social science explanations of the bimodality **8**
4.1 Discrete Markov transition model . . . 8
Model Details
4.2 Saturation of variables . . . 11
4.3 Undersampling . . . 11
5 Supplementary Note 5: Plots of PC1 by Time **12**
Supplementary References **20**

1 Supplementary Note 1: Overview of Seshat

The Seshat project was initiated in 2009 to develop "a general resource for testing theories about socio-cultural evolution and cliodynamics" [1, p. 59]. Here we provide an overview of the Seshat database including its structure (1.1) and variables (1.2), identifying some issues raised by the imputation procedure in [2]. Seshat is an evolving database, and so this summary largely refers to the version of Seshat used in [2]. In section 2 we briefly summarize two more recent Seshat publications. The remainder of this SI analyzes the source(s) of the bimodality in polity scores on PC1 (section 3) and for reference plots the PC1 scores through time for each NGA (section 5).

The Seshat database divides the world spatially into distinct "Natural Geographic Areas" (NGAs) and then temporally subdivides these by "polities" each of which lasts for one or more centuries. An NGA is a purely geographic unit, defined as an area roughly 100 km by 100 km "delimited by naturally occurring geographical features (for example, river basins, coastal plains, valleys, and islands)" [2, p. 6]. Examples of NGAs are the Cambodian Basin, the Konya Plain, and the Big Island of Hawaii. Each NGA is controlled by a sequence of polities, with at most one polity controlling the NGA at any time. A polity is "an independent political unit that ranges in scale from groups organized as independent local communities to territorially expansive, multi-ethnic empires" [2, p. 2]. Seshat uses a representative quasi-polity when more than one political entity existed in an NGA at one time, or when control of the polity switched rapidly among political entities – see[1] for further details. Following Seshat usage[2,3], we do not distinguish quasi-polities from polities in our analysis.

Generally, polities are defined more granularly than in colloquial speech, and a new polity is assigned when there is a significant sociopolitical realignment. For example, Seshat recognizes eight polities in the Kingdom of France from Charlemagne to Louis XIV. Seshat encodes measurements of societal complexity (where available) at 100-year intervals for each NGA and its associated controlling polity. The relationship between NGAs, centuries, and polities can be clarified by the example of Iceland, an NGA. There are four observations for Iceland, one every century from AD 1000-1300 inclusive. The first three observations (AD 1000-1200) are coded as a single polity (the Commonwealth), and the last (AD 1300) as a separate polity (part of the Danish Kingdom).

1.1 Choice of NGAs

A subset of NGAs in Seshat was used in the analysis of Turchin and colleagues[2]. This set was chosen to maximize geographic extent and diversity in social organization. In particular, 30 NGAs were sampled from 10 world regions (e.g., South America; Africa). From each world region three NGAs were selected to provide one example each of a place where complexity appeared early, intermediate, and late in date. The resulting dataset contains 414 distinct observations across NGAs at 100-year intervals. Seshat is an evolving dataset with the versions resulting in publications available for download at http://seshatdatabank.info/datasets/.

1.2 Choice of Variables and Byproducts of Imputation

The Seshat database now contains over 1500 variables. However, only a subset was used for recent articles analyzing complexity (and by us). Turchin and colleagues[2] chose 51 variables related to social complexity that they believe can be reliably coded even for (well-known) archaeological contexts. These variables are imputed into a set of nine distinct summary variables, called "complexity characteristics" (CCs). The nine CCs measure: (1) polity population, (2) extent of polity territory, (3) population size of largest urban center, (4) hierarchical complexity (administrative, religious, military, settlement), (5) government, (6) infrastructure, (7) information systems, (8) specialized literature, and (9) monetary system (see also Table 1, main text).

A problem with this imputation is that there are frequently one or more of the original variables for which values are missing. This can be due to incomplete research, expert disagreement, or other uncertainties about the correct value. Missing values are especially common for earlier polities for which less information is available. Rather than assign a value and create one canonical imputation, instead a distribution of possible values is assigned. These distributions are then sampled 20 times, to produce 20 replicates of the dataset [2, p. 7]. The imputation is performed on each replicate, producing 20 different sets of CC values. These replicates are used to produce confidence intervals on the proportion of variance explained by each PC, the component loadings, and the values of the PCs for each polity. The authors graciously provided the 20 imputed datasets, forming the basis of our analysis here. The complete dataset contains 8280 rows (414 observations $*$ 20 imputations), and 13 columns. The columns are: the name of the NGA, an identifier for the polity, the century of the observation, the value for each of the nine CCs (the result of imputation), and the imputation replicate identifier.

It is visually apparent that there are artifacts in the data shown in Supplementary Figure 2. The data exhibit noticeable "streaks" at approximately a thirty-degree angle from the vertical. We examined the cause of these streaks to ensure they are not a significant contributor to our findings. We first examined whether the streaks are caused by strong correlations among (a set of) CCs. We examined this by reproducing Supplementary Figure 2 but zeroing out the contribution from all combinations of 1, 2, and 3 CCs.

The character of the streaks (primarily the orientation) changes when certain CC combinations are

removed, but all subsets still contain visible streaks. This implies the streaks are not caused by strong CC correlations. It also implies they are not caused by the fact that some CCs are more "discrete" than others, for instance measured in integer increments from 1 to 5 rather than a more continuous measure. Instead, these streak artifacts are caused by the procedure used by[2] to generate the 20 dataset replicates. Each streak is, in fact, a single polity with its 20 replicates. The apparent directionality of the streaks arises from the uneven uncertainty assigned to each CC. Population measures, for instance, are specified in a range that is sampled, while Money Level CC was not assigned an uncertainty. In between are multiple uncertainties that are averaged into a single CC, and so their combined variance is reduced compared with population. These streaks give a visual representation of the effective uncertainty assigned to each polity. We find that the measured uncertainty in the Seshat data does not impact the results of the analysis in this work, nor the analysis of Turchin and colleagues[2]. The variation in each polity is small compared with the variation between polities. This level of measured uncertainty does not affect the existence of two strong clusters, and this noise, if anything, reduces the strength of the primary principal component rather than enhancing it. We incorporate the uncertainty from the different imputations in reporting the rest of our results.

2 Supplementary Note 2: Previous Seshat Research

We do not attempt a comprehensive review of the rapidly growing set of articles from this project. Instead, we summarize three recent, relevant articles. The foundational article in this line of inquiry[2] describes and analyzes the dataset we consider here. The major finding is the primary principal component (PC) captures 77% of the variation in the "complexity components" (CCs), with approximately equal loading by all nine CCs onto that first PC. Small-scale polities (typically earlier in time) have generally low scores when projected onto this primary PC, while larger-scale polities (typically later in time) generally have a high projection (score). The location of any single polity along this one, nine-dimensional PC might plausibly serve as a useful scalar measure of how complex, or possibly "developed", that polity was, in the sense of sharing features with large-scale rather than small-scale societies.

2.1 Evolution of Information Systems

The time-stamping of each observation is used profitably in two other recent papers from the Seshat group. The first[4] uses multiple linear regression to explain variation through time in the information complexity CC (*Info*). Potential independent variables lead the dependent by one century so that causal inferences respect time's arrow. There is also an autoregressive term for the dependent, to examine the extent to which earlier values for *Info* predict the current value. Finally, there are terms examining correlations due to shared culture history (proxied as language family) and spatial diffusion. The best linear model (that with the smallest AIC) identified first- and second-order lags of *Info* to be the most important predictors of *Info*, followed by phylogeny, *Money*, and *PolPop* (polity population), all with positive slopes. *Money* is a scale reflecting the "most sophisticated" monetary instrument present in the coded polity (0: none, 1: Articles, 2: Tokens, 3: Precious metals, 4: Foreign coins, 5: Indigenous coins, 6: Paper currency) [4, p. 39]. Broadly similar results were achieved when quadratic (squared) terms for the regressors were introduced although the Lag1 squared term for the autoregressive version of the independent variable bore a negative sign, possibly suggesting a "regulatory" effect for this term: "when *Info* gets too high, this term "steps on the brakes" to impede its further increase" [4, p. 51].

2.2 Evolution of Specialized Governance Structures

Another recent paper[5] employs the same approach to analyze the evolution of the internal specialization of governance structures (*Gov*) which sums 11 binary variables reflecting the presence/absence of various specific offices, government buildings, examination systems, etc. [5, p. 3]. Here, the lag1 version of *Gov* is by far the most important variable in the nonlinear regression, followed by the phylogeny proxy, a quadratic version of *Gov*, and (log-transformed) polity population (all signs positive except for *Gov*). The only other independent variable with a significant *t* value is *Time*, the absolute date of each time step. The authors interpret the inclusion of this variable in all the best-fitting models to indicate that important unobserved variables remain to be discovered and added to the model.

As valuable as these two analyses are, particularly since they take time explicitly into account, they deal with just one (dependent) variable at a time. Our re-analysis of the data in the main paper[2] where all the variables are considered simultaneously, is designed to see what gains can be made when that analysis is performed in a dynamic mode.

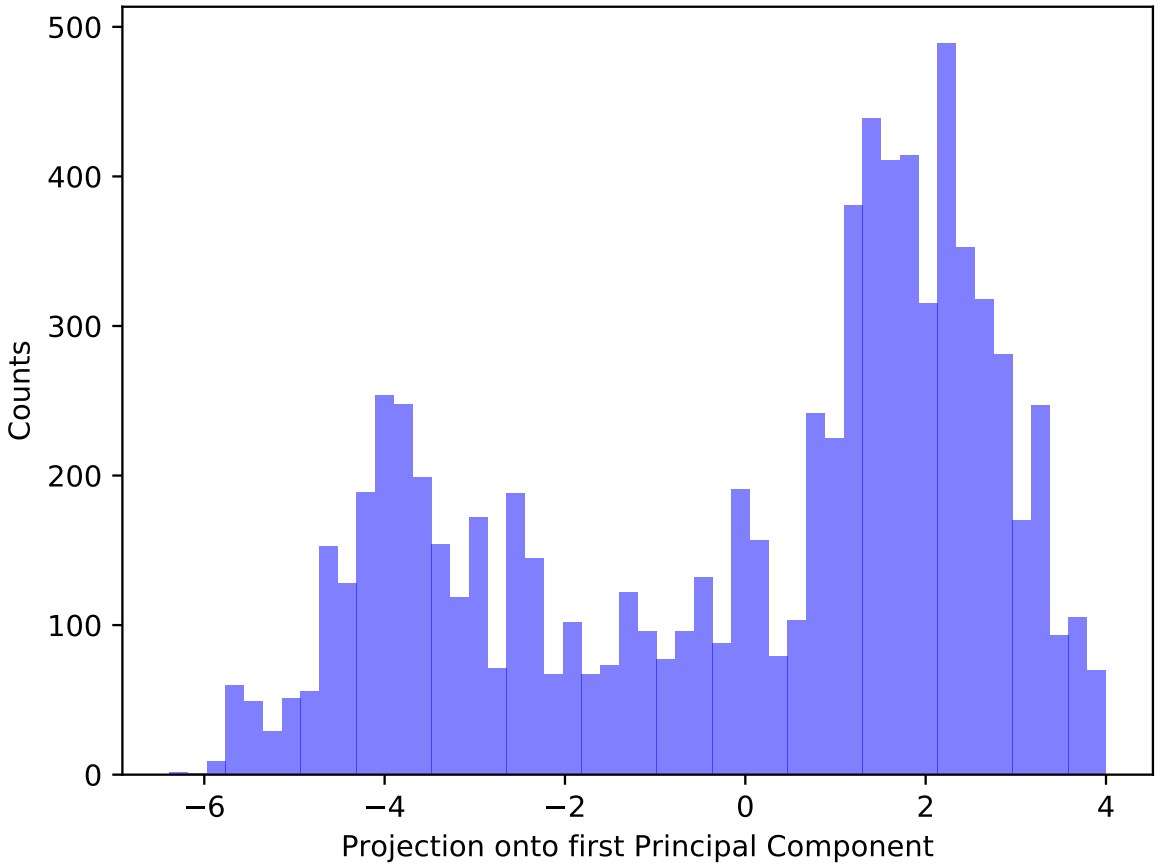
3 Supplementary Note 3: Bimodality

Our reanalysis of the Seshat dataset reported by Turchin[2] was originally motivated by the apparent bimodality in the frequency of polities projected onto their PC1 as illustrated in Supplementary Figure 1. This was first of all a concern: the primary formal justification for an analysis based on PCA (as in[2]) is the assumption that the data are generated by IID sampling a *single* Gaussian. The presence in these data of two highly distinct clusters is therefore problematic. Indeed, as shown in Supplementary Figure 1, the clusters in PC1 appear very similar, being accurately described by a mixture of *two* Gaussians of very similar widths, distinguished only by their mean values on PC1. Other results emerging from the PCA, such as the finding that PC1 captures "77% of the variability" become unreliable given this data structure. That is one justification for our decision to pay more attention to PC2, since, likewise, it may be more important than the statistics generated in[2] suggest. Accordingly, as a conservative first step, in this paper we consider the Seshat dataset projected onto the top two PCs.

The second reason the bimodality is of interest however is that — if not an artifact of sampling or analysis — it may suggest a fundamental division of societies, perhaps into "simple" and "complex" groups, that additionally might be regarded as fundamental basins of attraction for the dynamics of social groups as they move through the Holocene. (For example, on the basis of a somewhat similar analysis Peter Peregrine[6] has recently argued for the existence of four modes in sociopolitical organization.) One of these clusters (with a modal score around -4 on PC1) corresponds to polities of modest demographic scale, often but not invariably early in time. The other cluster (with a modal score around 2 on PC1) corresponds to polities of larger demographic scale, generally appearing later in time. As a shorthand we will refer to these as the "early" and "late" clusters, though this is not invariably true in calendar years. The only polities that reach the scale required for stable membership in the late cluster, at least in this sample, are from the Old World (Eurasia, Africa, and Oceania).

Turchin and colleagues[2] also recognized these two clusters. On pp. 26-28 of the supplemental information they note "some initial support for the idea" of clusters in the distribution of societies along PC1 "with a relatively large distance between two main clusters". Figure SI11 in that paper suggests that two modes begin to develop along the PC1 dimension by 1000 BC and are quite prominent by AD 1. Similarly, Turchin et al. [5, p. 13] report that "there may be two clusters in the governance–population phase space, where trajectories tend to spend most time. The transition between the two clusters, on the other hand, happens quickly."

We now present several analyses that suggest that the bimodal distribution of points along PC1 is



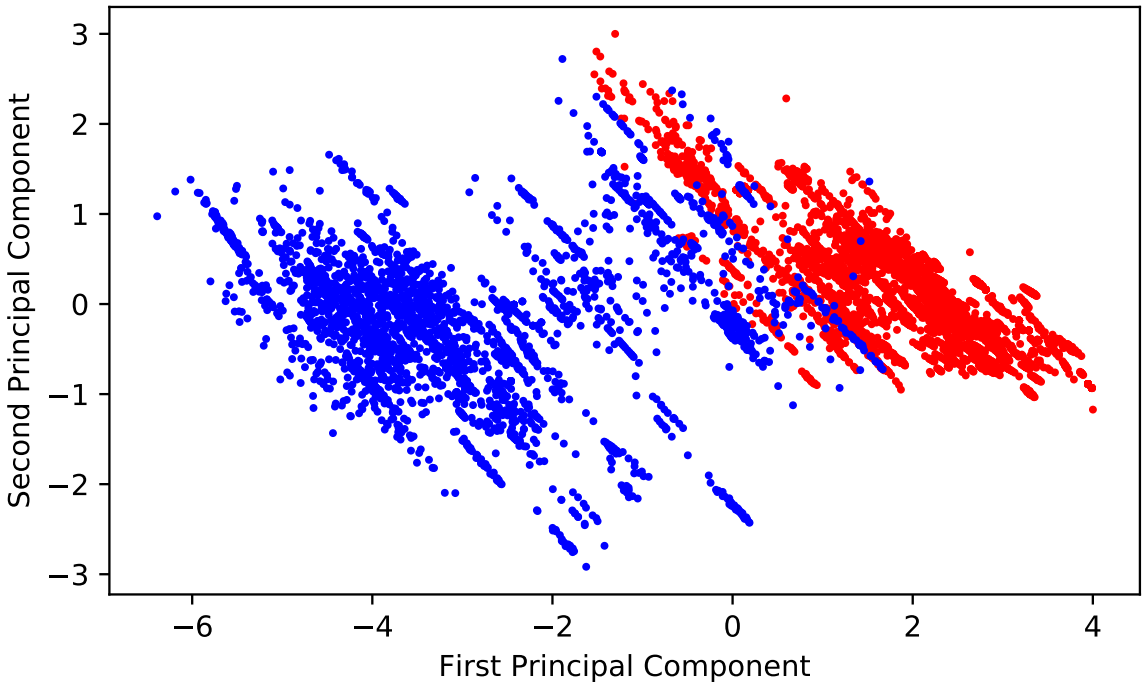
Supplementary Figure 1. Histogram of polities projected onto PC1. Each imputation comprises a separate point in the underlying data for the histogram.

likely not of substantive interest, though this structure still raises issues for the strict interpretation of the proportions of variance explained by the various PCs. In particular, the bimodality is *not* the result of the streak artifacts discussed in 1.2. We also show that the bimodality is not the result of saturation of the variables recorded in the Seshat dataset at their extremal values as societies develop. (Such saturation may however contribute to the accumulation of datapoints in the right-hand cluster). There were also some possibilities we could not reject. In particular, we could not conclude that there is not an attractor of the underlying dynamics (i.e., a "sink") in the right-hand cluster.

3.1 Gaussian Mixture Model

To investigate the bimodality in more detail we began by fitting a two-component Gaussian mixture model to the full 9-dimensional (9D) imputed Seshat dataset. Supplementary Figure 2 presents a 2D scatterplot of PC1-PC2, with the coloring of the points determined by which 9D Gaussian cluster each point most likely belongs to. We define cluster membership for each point as its maximum likelihood cluster, i.e., the Gaussian with the higher membership probability. (Note that the true data are 9-dimensional, and the full data are used to define membership, given the seemingly non-monotonic clustering classification.)

Moreover, as shown in Supplementary Figure 3, most of the data points are included in the $< 99\%$

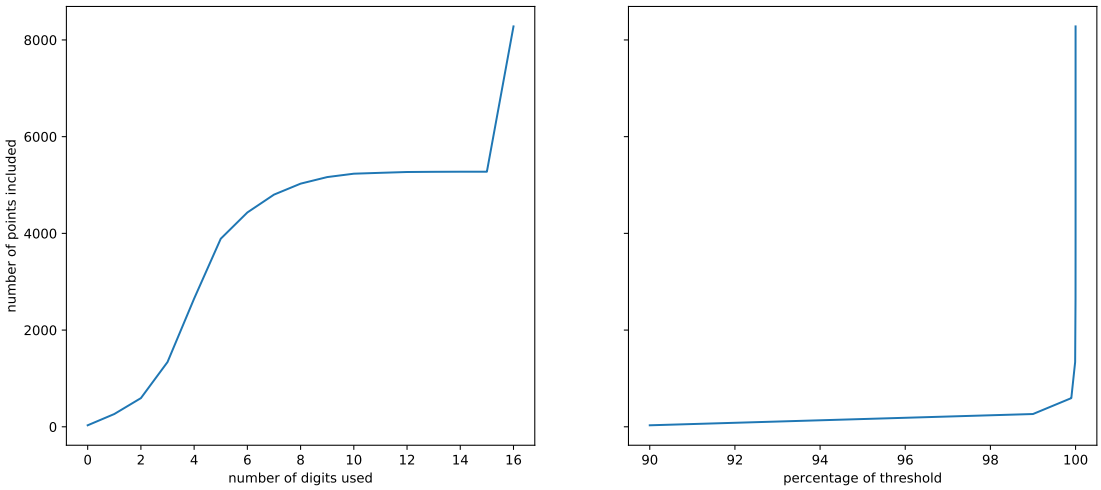


Supplementary Figure 2. Scatterplot of polities projected onto the first and second PC. The left (blue) and right (red) clusters are generated by fitting a two-component Gaussian mixture to the 9-dimensional PC data and determining, based on the mixture densities, whether a point belongs to the first or second cluster. The overlap between points in the cluster occurs because the 9-dimensional fit is displayed projected into the 2-dimensional PC1-PC2 space. In fact, a point belongs exclusively to one of the two clusters as suggested by Supplementary Figure 3. Note the pronounced "streaking" from top left to bottom right which appears to be due to the imputation procedure.

threshold, meaning that most points exclusively belong to one of the two clusters according to our Gaussian Mixture Model. This result suggests that the two clusters we observed in the data are clearly distinguishable.

To test whether the two Gaussian clusters were due simply to noise, we bootstrapped the following quantities: 1) the percentage of the sum of all eigenvalues that is given by the largest eigenvalue alone, in the first and the second Gaussians, 2) the angle between the principal axes of the two Gaussians, corresponding to the largest eigenvalue in each Gaussian, and 3) the angle between PC1 and the principal axis of the first Gaussian, and the angle between PC1 and the principal axis of the second Gaussian. Note that all these values were obtained by fitting two Gaussians in the 9-dimensional imputed Seshat data matrix.

In more detail, the these properties of the Gaussians were calculated as follows. 1) To compute



Supplementary Figure 3. Result for verifying the robustness of the two Gaussian clusters. After fitting the Gaussian mixture model with two components, using the predicted posterior probability for each component for each data point, we determine how many points have above the given threshold probability of belonging exclusively to one of the two components. Left: total number of points included in either of the clusters based on the number of digits used for the threshold probability (e.g., 6 digits for the threshold means that the threshold was 0.999999). Note that in this left-hand plot, when the number of digits used for the threshold reached 16, every data point was included in the clusters due to numerical overflow error. Right: number of points included in the two clusters when changing the threshold of the probability value (quantified as a percentage) for belonging exclusively to one or the other of the two clusters (e.g., "90 percent threshold" picks out data points that have less than 90% probability of belonging exclusively to either of the two clusters).

the percentage of the largest eigenvalue in the two Gaussians, we first calculated 9 eigenvalues for the covariance matrix of each of the two fitted Gaussians. For each Gaussian, we then selected the eigenvalue with highest value and divided it by the sum of all 9 eigenvalues to obtain the percentage of the largest eigenvalue. 2) To compute the angle between the two main eigenvectors, we selected the eigenvector that corresponded to the largest eigenvalue in each of the two Gaussians. Then, we calculated the angle between the two vectors using the cosine formula. 3) Computing the angle between Turchin's main principal axis and the main eigenvectors of the first and the second Gaussian follows exactly the same procedure as above, by taking the main principal axis vector and computing the angle between that vector with the main eigenvectors of the two Gaussians.

The procedure for bootstrapping was as follows. First, for each of the methods listed, we randomly resampled the original 9-dimensional imputed data matrix and obtained 5000 matrices that are of the same dimension as the original matrix. Then, for each of these 5000 matrices, we computed the statistic of our interest (e.g., angle between the two main eigenvectors) and obtained an array of resulting values for the statistic. We plotted the values calculated from resampled matrices, as well as plotting a vertical line that indicates the value calculated from the original matrix. Furthermore, we indicated the upper and lower bounds of the bootstrap normal 95%-confidence interval with dashed lines, which were computed as follows: suppose that for an estimate T of interest, bootstrap estimates are given by $T_1, ..., T_n$ with original sample estimate T_0. Then, the normal 95%-confidence interval is given by T_0 - bias $\pm$ $std(T_1, ..., T_n)$,

where bias is defined as $mean(T_1, ..., T_m) - T_0$ and $std(T_1, ..., T_n)$ refers to the standard deviation of the bootstrapped estimates. The result (Supplementary Figure 4) suggests that the outcome is robust, implying that the two Gaussians were not caused by noise.

So we wish to ask whether this the bimodality is of substantive interest, or some kind of an artifact of the statistical analysis or the data-generation methods. (Note that some of those data-gathering methods are implicit, arising in how archaeologists choose where to excavate and what variables to pay attention to. Others are explicit, for example the decision by the creators of Seshat to exclude many polities from the database to try to get uniform coverage over NGAs, or of the use of imputations to "fill in" holes in that database.) We examined several candidate hypotheses for the bimodality and were able to statistically reject several of them — but not all. Some of these null hypotheses are discussed in the next section.

4 Supplementary Note 4: Possible non-social science explanations of the bimodality

Imagine for a moment that the time-stamping of all the polities were randomly re-assigned while keeping all the other attributes for each polity intact. A PCA would still "explain" 77% of the (static) patterning in the data, but each polity's movement along the continuum defined by PC1, over time, would be random. In that case, would we think that PC1 provides a useful summary of sociopolitical development? Probably not, since we understand (sociopolitical) evolution as an historical process. What really counts for understanding such a process is not the "static pattern" formed by superimposing a sequence of snapshots of the process. Rather we ought to be interested in the dynamics of that entire sequence. This observation alone warrants examination of the dynamics of the data, ideally in more dimensions than just PC1. Indeed, as described below, many aspects of static patterns can be misleading, leading to conclusions that in fact are contravened if one directly analyzes the dynamics. This issue is not assessed by[2] (except for noting after having performed the PCA that early polities tend to have lower values of PC1 than later polities).

4.1 Discrete Markov transition model

To more directly address whether the dataset reflects variation in the underlying dynamics as one moves across PC1 or just a static pattern, we constructed a novel null-hypothesis technique. This technique is based on the observation that even if the underlying dynamics through PC space is a time-homogeneous Markov chain, with no basins of attraction, and no dependence on time, the *dwell times* in the various parts of PC space caused by evolving under that Markov chain could vary significantly. That heterogeneity in dwell times might in turn cause the observed bimodality, ironically due to the fact that datapoints are sampled uniformly in time. In essence, the uniformity of sampling in time would introduce an artifact of *non*-uniformity of the distribution over space.

To investigate this hypothesis, we binned PC1 and then used the (time-stamped) Seshat data to estimate the conditional distribution of moving from one bin to another over a given (fixed) time interval, Δt. We then iteratively applied that conditional distribution to every datapoint in Seshat, *exactly* $t/\Delta t$ *times* (where t is the number of years ago of the timestamp of that datapoint). We then compared the resultant distribution of points to the full Seshat dataset. This amounts to assessing the null hypothesis that the distribution of points in PC1 arises from a combination of the following phenomena:

a) Random variation across the NGAs in the PC1 value of the earliest polity within that NGA;

b) Random variation across the NGAs in the chronological time of that first PC1 value;

c) A general drift of polities from low to high PC1 values driven by a time-homogeneous Markov chain.

Our tentative conclusion is that the data are consistent with our null hypothesis, i.e., that these three phenomena, by themselves, can explain the bimodality in PC1. Moreover, the conditional distribution of next bin given current bin under the estimated Markov chain of item (c) is irreducible (by inspection — see below). So there is a unique fixed point of this null hypothesis dynamics. This indicates that there is no evidence for there being more than one "basin of attraction" of the dynamics.

In addition we performed a very preliminary analysis of the relationship between the precise details of a stochastic process traversing a space and the resultant static pattern given by sampling that process at regular time points (as was done in generating Seshat, where the sampling time points were one century apart). Obviously, if the expected speed of movement across PC1 is slow for a broad interval of PC1, then is fast in a following interval, and then slow again in a following, third broad interval, there would be an increased number of datapoints in that first broad interval as well as the third one, with few points in between (given that the datapoints are formed by sampling at times all one century apart). This would result in the kind of bimodality in PC1 we found in the Seshat dataset. Conversely, if mean speed across PC1 were uniform across PC1, *and all other aspects of the dynamics were also uniform*, there would be no such induced bimodality in the samples of PC1 values.

However, we realized that even if the mean speed is uniform across PC1, variation in the *breadth* of possible speeds from one point in PC1 to another can result in bimodality of the sort that characterizes the Seshat dataset. To see why, first note that the amount of time a given polity will spend in a given interval of PC1 as it moves across PC1 will depend on the *inverse* of the speed across that region. Moreover, if we have two distributions over speed values that are both Gaussian with the same mean, but one of those distributions has a bigger variance than the other, the associated expected *inverse* speed will be greater in the distribution with a larger variance of speeds. This means that there will be a greater density of points sampled from a region of the space with a high variance of speeds than from a region of the space with a low variance of speeds.[1]

Accordingly, as a variant of our null-hypothesis test of what explains the bimodality, we modified it to try to correct for this dependence of dwell times on the variance of speeds. We did this by modifying how we estimate the conditional distribution of the Markov chain, to force no such variation across PC1 in the variance of the velocity from one point in PC1 to another. More precisely, we first estimated the average over all the bins of the variance of the speeds leaving that bin, σ^2. Then for each column probability vector p in the conditional distribution matrix of next bin given current bin, we solved for the distribution q with minimal Kullback-Leibler divergence to p which has its variance of speeds equal to the global average of such variances, σ^2. We then replaced p by q. Doing this for all columns in the original conditional distribution matrix gives a new conditional distribution matrix, which we could then apply to the starting polity position / times of the datapoints in Seshat, just as before. This variant of the null hypothesis appears to give a slightly worse fit to the actual full Seshat data than the original null hypothesis, in which the variances of speeds are allowed to vary across PC1. This leads us to conclude that there is weak but inconclusive evidence that variation in the variance of the speeds of movement across PC1 contributes to the bimodality.

The details of these analyses are presented in the next subsection. Before presenting them though, it is important to emphasize that while the *static* bimodal pattern can be (mostly) explained as a combination of phenomena (a)-(c), the ultimate question for us is what patterns there are in the *dynamics* across the space. To a degree, this can be assessed by examining the Markov transition matrices made under our null

[1] As a simple illustration, if the speed of movement across a bin of width 1 is 1, with probability 1, then it will take 1 unit of time to traverse that bin. If instead the speed when traversing that bin is either 1.5 or .5 with equal probability, the mean speed is again 1, but it will take on average $(2+2/3)/2 = 4/3$ units of time to traverse the bin. So there would be a greater expected dwell time in that bin.

models. However, an important direction for future work is to use more sophisticated statistical techniques to tease out that dynamics from the limited data.

4.1.1 Model Details

In light of the extremely small amount of data, we considered only the dynamics over PC1, and discretized those PC1 values into 10 bins. We then formalized the phenomena (a)-(c) in 4.1 with the following model:

1. NGAs are "born" (i.e., first possess polities of sufficient levels of social development to warrant inclusion in the Seshat database) by IID sampling a distribution p_0 over those 10 bins;

2. The chronological time of the birth of each NGA is given by IID sampling a distribution $\Delta(t_0)$;

3. All NGAs evolve across the bins over time-jumps of 500 years by sampling the same (time-homogeneous) conditional distribution π;

4. All data points, over all times, are subject to observational noise (including random removal from the dataset entirely) to generate the final dataset.

For simplicity, we estimated π by applying a frequency-count estimator to the 500-year bin-to-bin transitions in Seshat (shown in Supplementary Figure 5). However, to be maximally conservative, we avoided estimating $\Delta(.)$ and $p_0(.)$, by pursuing the following procedure instead: First, for each NGA n, we found the earliest bin at which it is present in Seshat, $i(n)$. Let $T(n)$ be the number of 500-year blocks into the past of that earliest data point $i(n)$. We applied π to $i(n)$ a total of $T(n)$ times, to generate a set of $T(n)$ bin values; doing this many times gave us a scatterplot of the PC1 values that we would expect to see for that NGA, if indeed its dynamics was given by π. Running this procedure for all of the NGAs gives an overall scatterplot which, if our model is correct, should look similar to the actual Seshat scatterplot. Supplementary Figure 6 (A) shows the results from this sampling procedure, which clearly exhibits bimodality similar to the original data.

We used this simple model to investigate the dynamic cause of the bimodality. Recall that the Seshat datapoints were recorded for regularly sampled times one century apart. This means that the two clusters would arise if movement through PC1-PC2 space was smaller where the first and second clusters lie, and faster in between them. In other words, it would arise if there were heterogeneity in the mean velocities across PC1-PC2 space. In terms of our discussion just above, this would correspond to variation, across the rows of the transition matrix π, of the difference between the average column index specified by the distribution of that row and the row index itself, i.e., variation with bin n of the difference $\left[\sum_m \pi(m \mid n)m\right] - n$.

Perhaps more surprisingly, *even if the mean velocities do not vary*, if the *variance* in the mean velocities varies in the appropriate way, then clusters would arise in regions where that variance is relatively large. To see this, note that the amount of time an NGA stays in a given PC1 interval is proportional to its *inverse* speed across PC1 in that interval. Therefore a Gaussian distribution in speeds will result in a non-Gaussian distribution in the amount of time an NGA stays in a given PC1 interval, i.e., a non-Gaussian distribution in the number of Seshat data points in that interval.

To test these possibilities, we first generated a new transition matrix that resembles the original matrix π, except the mean and/or variance of speeds across the bins was forced to be the same, no matter what the originating bin. For each bin n, we generated the associated conditional distribution (i.e., row) of this alternative transition matrix, $\pi'(.|n)$, by minimizing the KL-divergence between $\pi'(.|n)$ and $\pi(.|n)$, subject to the constraints that the mean and/or variance of that new conditional distribution equals the associated value for the entire population, running over all n. Supplementary Figure 6 B, C, and D simulate

the histogram with constrained mean velocity, constrained variance in velocity, and both, respectively. The bimodality observed in the data (Supplementary Figure 6 A) is still present in all these graphs. This suggests that variation in the mean and/or variance of speed in different PC1 regions cannot be the sole explanation for the bimodality, though we do not exclude the possibility that such variations contribute.

In addition to these possible explanations of the bimodality grounded in dynamics through PC1-PC2 space, it is also possible that the bimodality is simply caused by the sampling process that generated the initial distribution. That is, in Seshat the initial observation of NGAs is biased towards low PC1 values, which by itself would create the early cluster, even without any superimposed dynamic process. To see this, we simulated the histogram by setting p_0 to be a uniform distribution i.e. equal probability across bins, while keeping the transition matrix π the same as estimated. In Supplementary Figure 6 E, we see that the size of the left cluster is reduced significantly from what is actually found in the Seshat data. This implies that non-uniformity of the initial values of PC1 for each NGA explains the left cluster, at least in part.

4.2 Saturation of variables

There are several ways in which the observed bimodality could result from a statistical artifact in the sampling or definition of the data or variables. One scenario is that the two peaks are generated from saturation of the variables (the nine CCs). If the saturation of different variables occurs at different rates, then peaks would be observed around the time when each variable reaches its upper limit and the growth of PC is temporarily slowed.

To test for this, we tried some simple simulations where the PC is a sum of two variables, both of which follow logistic growth $(L_i/(1.+\exp(-k_i*(x-x_i^0))), i=1,2)$ but with different parameters. Supplementary Figure 7 (a)-(c) shows how the two variables with different saturation times sum up. Assuming polities that follow this growth path are sampled uniformly, the resulting histogram is shown in Supplementary Figure 7 (d). While the two peaks can be generated from this logic, the right peak is at the right-most time period, in contrast to the data graphed in Supplementary Figure 1. It is also worth noting that it requires us to fine-tune parameters to create such peaks. Supplementary Figure 8 shows different combinations of saturation parameters. In general, the uniform sampling results in a plateau at the right rather than two spikes if both of the variables are saturated within the observed period.

4.3 Undersampling

Another possible explanation for the bimodality is selection bias in the data. For example, perhaps polities of intermediate complexity were undersampled for some reason, such as lack of archaeological excavation or reporting for these polities, which might yield a bimodal distribution. To check that the bimodality (especially the peaks) was not caused by undersampling issues, we ran the following tests. Although it may be far-fetched to claim that these tests were sufficient to confirm that undersampling was not a problem, bimodality survived after each of these tests (Supplementary Figure 9).

1. Interpolating missing time periods within each NGA: We linearly interpolate PC1 by NGA to fill out the missing periods.

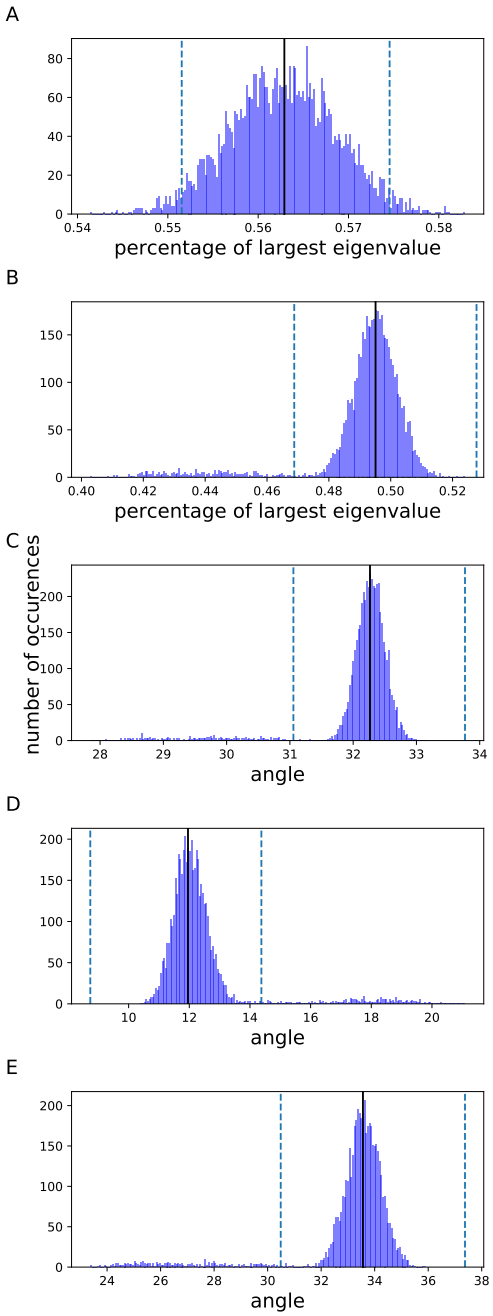
2. Weighting within each NGA depending on the sparsity of data within its neighborhood: For instance, say we observe an NGA at periods t_1, t_2, and t_3. We weigh the observation at t_2 by $\frac{(t_2-t_1)+(t_3-t_2)}{2}$ so that the observation in a sparse period counts more.

3. Use data points only from the NGAs that span both Gaussians: We take the NGAs that are classified as the first Gaussian in some periods, and the second Gaussian in other periods.

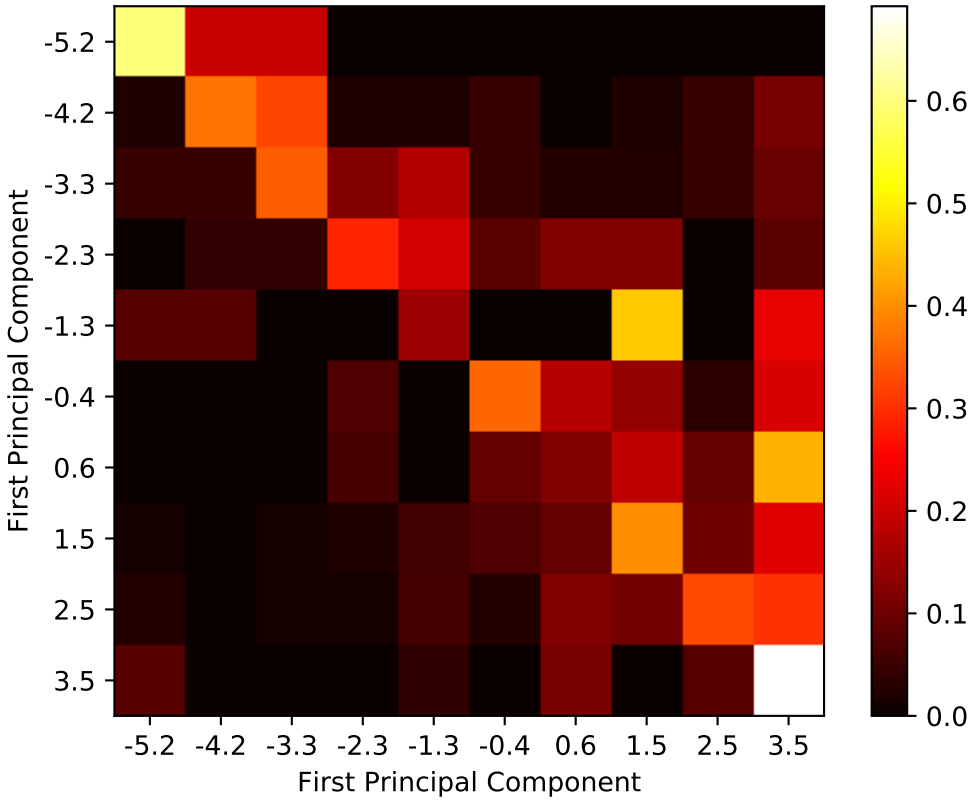
Supplementary Figure 9 plots the PC1 value by the normalized density for each of the scnearios above, linearly interpolating the main PC values for each NGA for missing periods; the two peaks remain for all these cases. We found the two clusters of points, each corresponding to peaks in the bimodal curve after doing a scatter plot under each method, with time and the main PC value as the two axes (not shown). Also, we found that there were fast transitions from one (lower) cluster to the other (higher cluster).

5 Supplementary Note 5: Plots of PC1 by Time

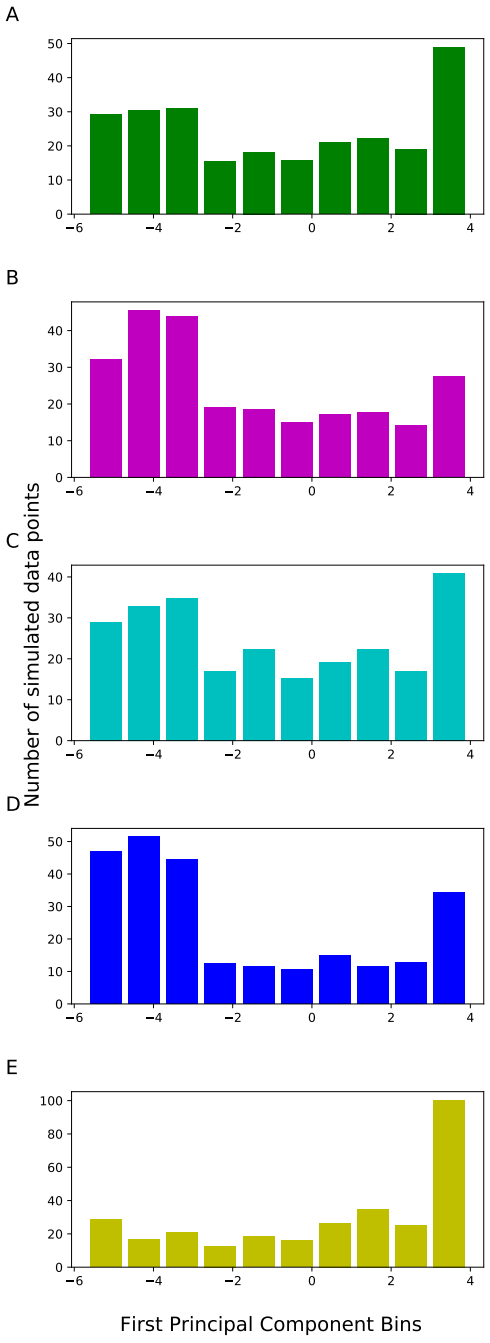
In the main text we provide several plots showing PC1 scores by PC2 scores. Supplementary Figure 10 provides plots of PC1 scores through time for each NGA. Note that if each of these 30 sequences exhibited a perfect positive correlation of its score on PC1 with its date, then we would be incorrect to maintain that the pattern-based correlational analysis undertaken by[2] cannot stand in for the dynamics. However, this is clearly not the case. Moreover, even if this were true, we see that these individual sequences exhibit very different slopes and intercepts — a fact of substantive interest though not pursued in this article. Of course we should be wary of putting too much emphasis on such a correlational analysis since some of the values in these figures are interpolated.



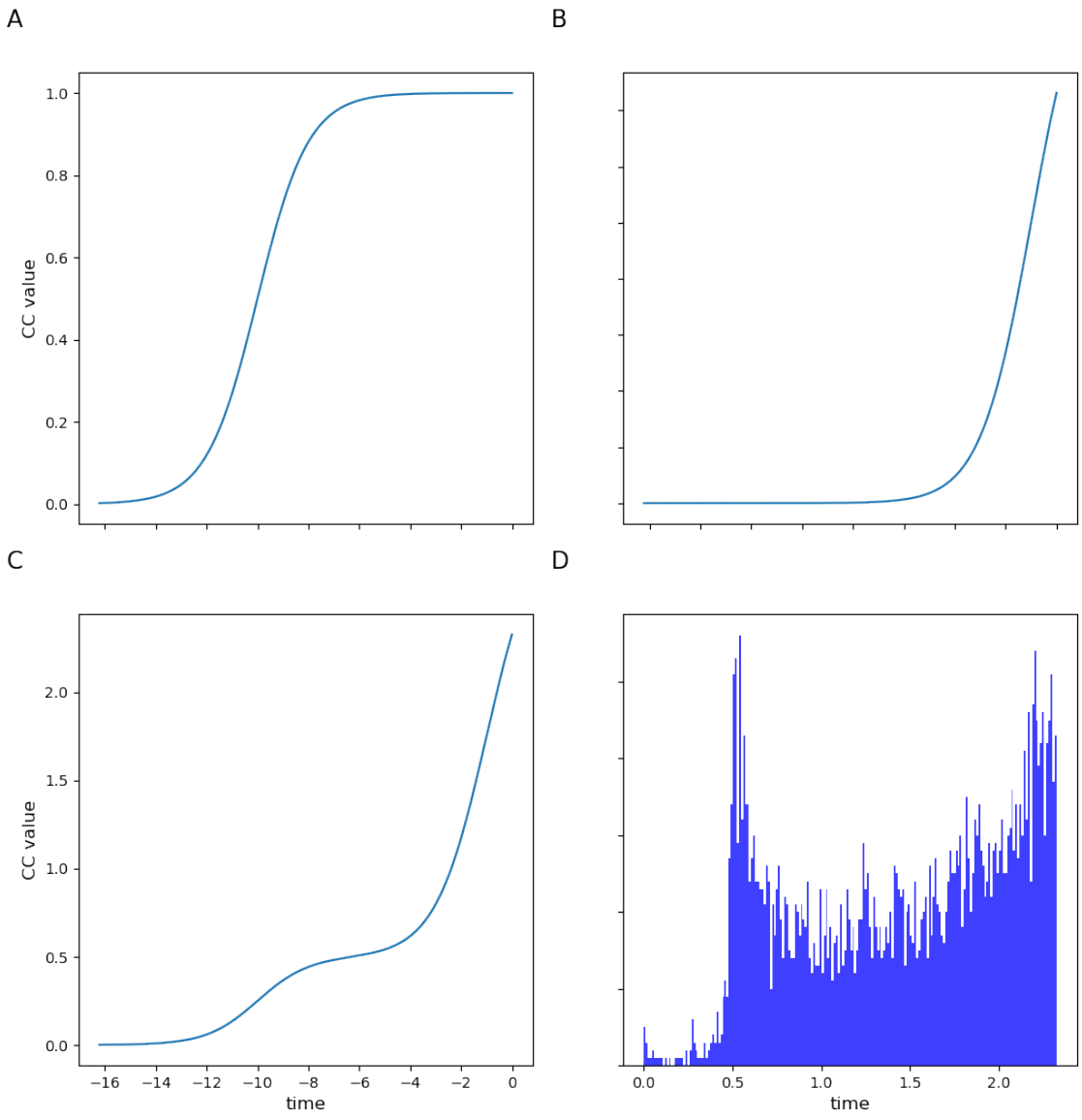
Supplementary Figure 4. Bootstrapping results for various statistics to verify the robustness of the two Gaussian clusters along with 95%-confidence interval. (A) The percentage of the sum of all eigenvalues in the first cluster given by the largest eigenvalue. (B) The percentage of the sum of all eigenvalues in the second cluster given by the largest eigenvalue. (C) The angle between PC1 and the principal axis of the first cluster. (D) The angle between PC1 and the principal axis of the second cluster. (E) The angle between the principal axes of the two clusters.



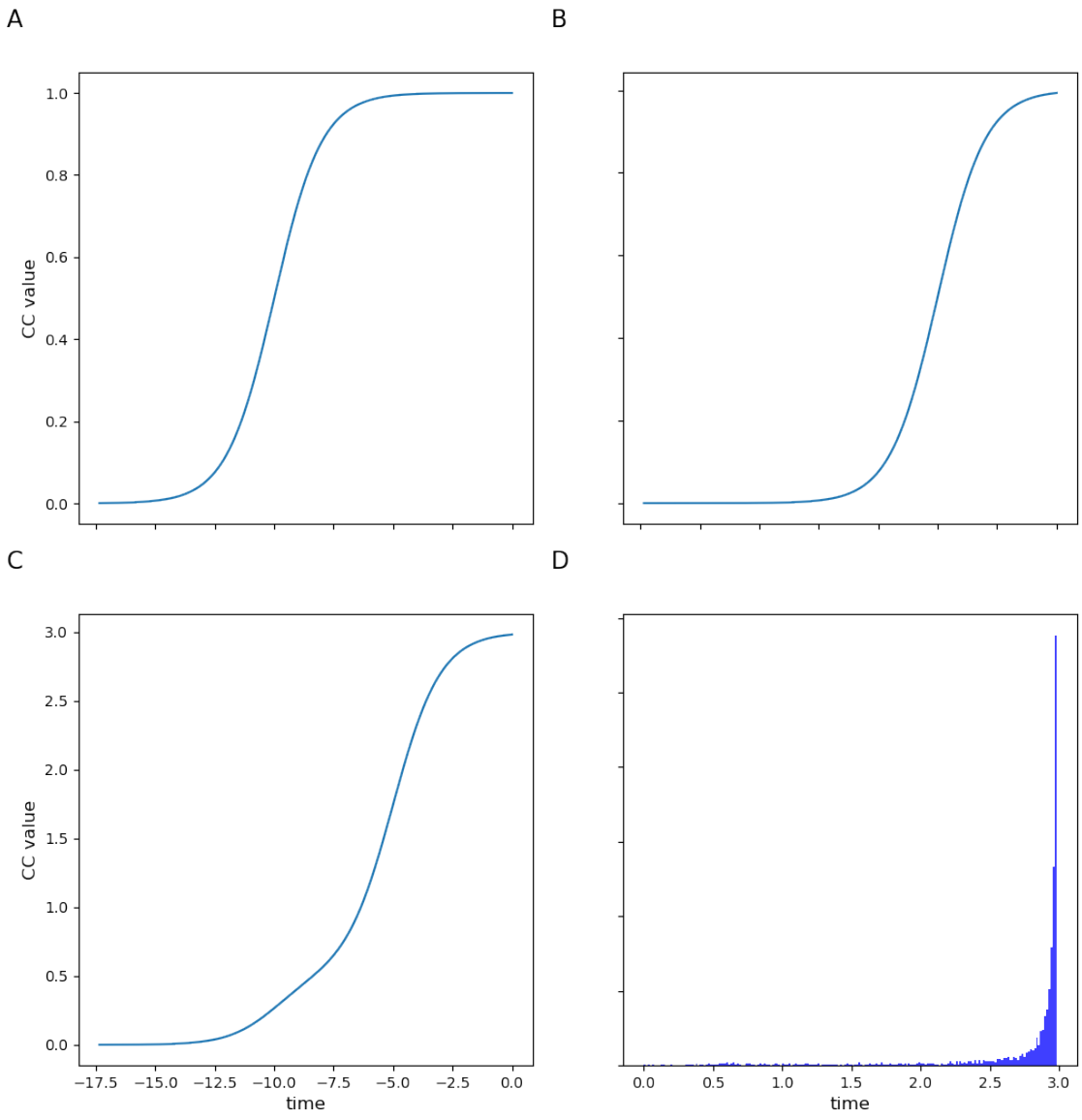
Supplementary Figure 5. The estimated transition probability π.



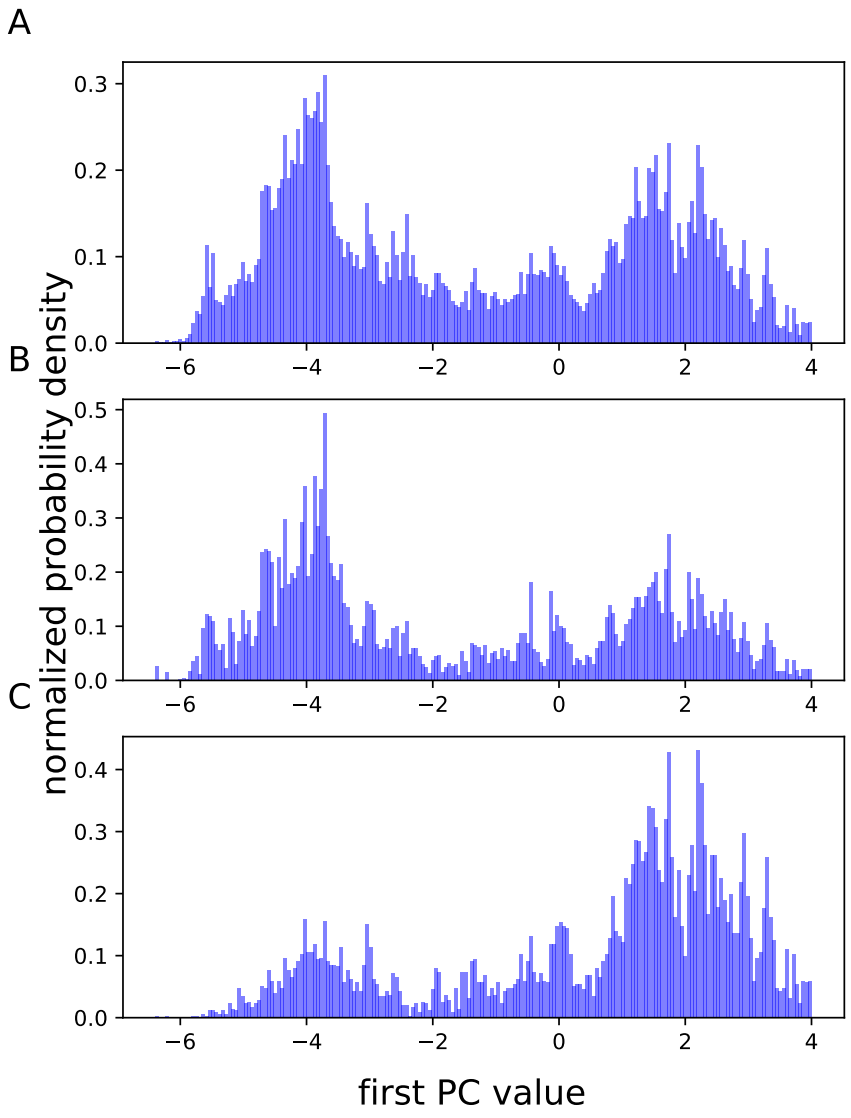
Supplementary Figure 6. Histograms of the simulated results from the Markov transition model. (A) The original data. (B) Constrained mean speed. (C) Constrained variance of speed. (D) Constrained mean and variance of speed. (E) Uniform distribution.



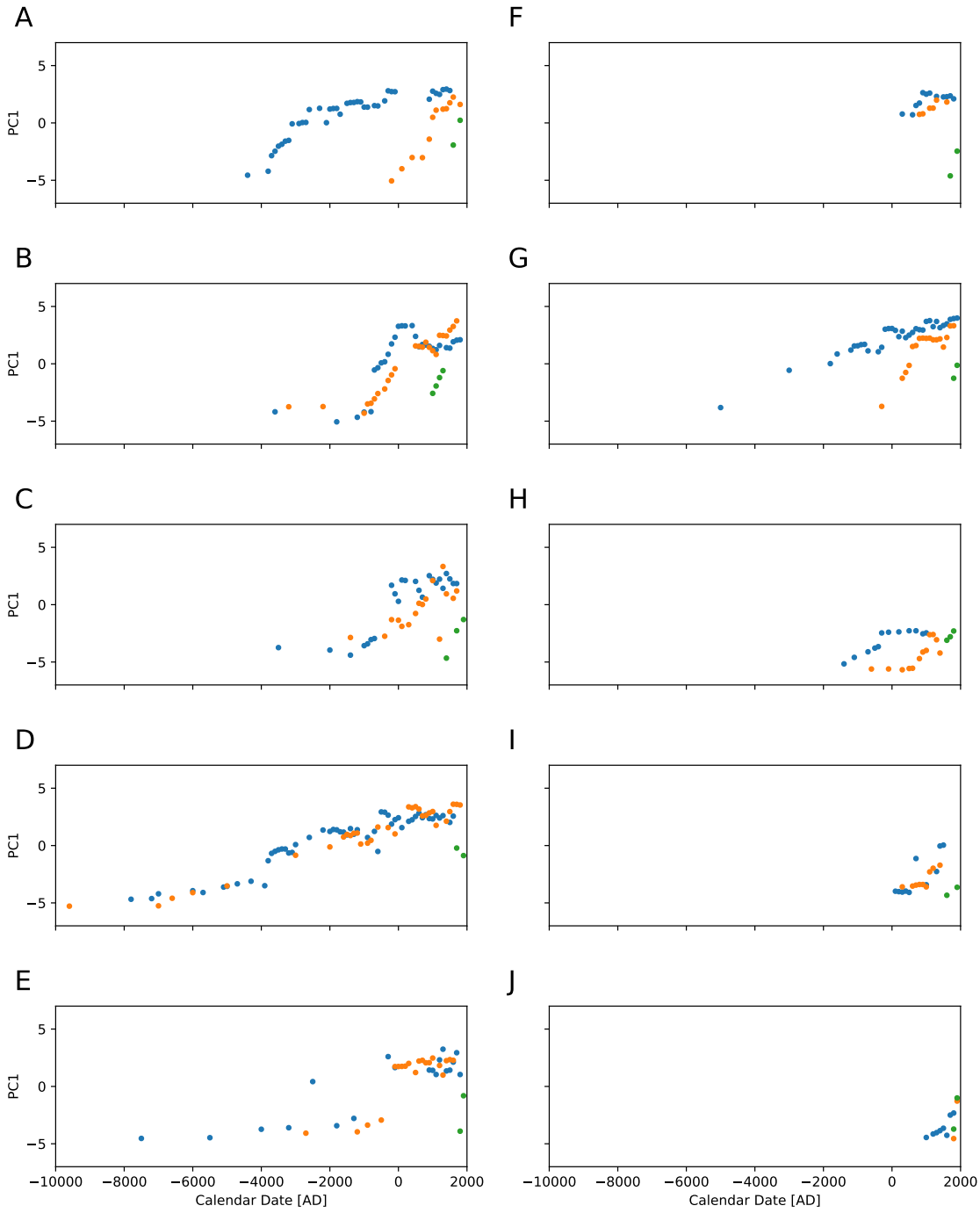
Supplementary Figure 7. The simulated logistic growth of two CCs. (A) First CC growth, (B) Second CC growth, (C) Growth of the average, (D) Histogram of the average under uniform distribution.



Supplementary Figure 8. The simulation of logistic growth CCs with different parameters. (A) First CC growth, (B) Second CC growth, (C) Growth of the average, (D) Histogram of the average of two CC.



Supplementary Figure 9. Checking for undersampling bias in the data through interpolation of time. (A) PC1 linearly interpolated for each NGA for missing periods. (B) Each observation weighted based on its relative sparsity with respect to each NGA, so that observations in sparse periods weigh more. (C) Only data points from NGAs that span both Gaussians



Supplementary Figure 10. Scores on PC1 as a function of time for 30 NGAs in 10 regions. Blue, orange, and green points are, respectively, NGAs in each region with early, intermediate, and late onsets of centralization. The regions are (left column): (A) Africa, (B) Europe, (C) Central Eurasia, (D) Southwest Asia, (E) South Asia; and (right column) (F) Southeast Asia, (G) East Asia, (H) North America, (I) South America, and (J) Oceania-Australia. The PC1 value is the mean across 20 imputations.

Supplementary References

1. Turchin, P. *et al.* Seshat: The global history databank. *Cliodynamics* **6**, 77–107 (2015).
2. Turchin, P. *et al.* Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proc. Natl. Acad. Sci.* **115**, E144–E151, DOI: 10.1073/pnas.1708800115 (2018). http://www.pnas.org/content/115/2/E144.full.pdf.
3. Whitehouse, H. *et al.* Complex societies precede moralizing gods throughout world history. *Nature* DOI: 10.1038/s41586-019-1043-4 (2019).
4. Turchin, P. Fitting Dynamic Regression Models to Seshat Data. *Cliodynamics* **9**, 25–58 (2018).
5. Turchin, P. *et al.* Evolutionary pathways to statehood: Old theories and new data. *SocArXiv Prepr. DOI: 10.31235/osf.io/h7tr6* (2018).
6. Peregrine, P. N. Toward a theory of recurrent social formations. In Sabloff, J. A. & Sabloff, P. L. W. (eds.) *The Emergence of Premodern States: New Perspectives on the Development of Complex Societies*, 271–295 (SFI Press, 2018).