DHQ: Digital Humanities Quarterly
2016
Volume 10 Number 4

A Macroscope for Global History: Seshat Global History Databank, a methodological overview

Pieter François <p_dot_francois_at_herts_dot_ac_dot_uk>, University of Hertfordshire, University of Oxford
J.G. Manning <joseph_dot_manning_at_yale_dot_edu>, Yale University
Harvey Whitehouse <harvey_dot_whitehouse_at_anthro_dot_ox_dot_ac_dot_uk>, University of Oxford
Rob Brennan <rob_dot_brennan_at_cs_dot_tcd_dot_ie>, Trinity College Dublin
Thomas Currie <T_dot_Currie_at_exeter_dot_ac_dot_uk>, University of Exeter, Penryn Campus
Kevin Feeney <kevin_dot_feeney_at_cs_dot_tcd_dot_ie>, Trinity College Dublin
Peter Turchin <peter_dot_turchin_at_uconn_dot_edu>, University of Connecticut

Abstract

This article introduces the "Seshat: Global History" project, the methodology it is based upon and its potential as a tool for historians and other humanists. Seshat is a comprehensive dataset covering human cultural evolution since the Neolithic. The article describes in detail how the Seshat methodology and platform can be used to tackle big questions that play out over long time scales whilst allowing users to drill down to the detail and place every single data point both in its historic and historiographical context. Seshat thus offers a platform underpinned by a rigorous methodology to actually do *longue durée* history and the article argues for the need for humanists and social scientists to engage with data driven *longue durée* history. The article argues that Seshat offers a much-needed infrastructure in which different skill sets and disciplines can come together to analyze the past using long timescales. In addition to highlighting the theoretical and methodological underpinnings, Seshat's potential is demonstrated using three case studies. Each of these case studies is centred around a set of longstanding questions and historiographical debates and it is argued that the introduction of a Seshat approach has the potential to radically alter our understanding of these questions.

Introduction

Seshat is the Ancient Egyptian deity of knowledge, writing and scribes. She is depicted working on a scroll but we will probably never know what she is writing. The few images we have of her show the scroll in profile, which is, from a viewers' perspective, razor thin. Of course, this does not stop us from imagining what Seshat is writing down. Given her twin tasks of keeping a record of the passing of time and surveying the land, we can well imagine the scroll might contain an endless variety of data, ranging from crop yields for individual plots of land to detailed lists of tax revenue arranged by name to the cost of building and maintaining temples. In short, we can imagine the scroll allowing us a means to measure the health of Ancient Egyptian society at any given point in time and link this with exact figures at a range of scales. This is of course a very modern imagination. Calls for tools that allow historical data to be queried equally well at different scales of analysis are proliferating within the world of digital humanities. Tim Hitchcock, for example, defined such "macroscopes" as: "a visual tool that allows a single data point, to be both visualised at scale in a context of a billion other data points, and drilled down to its smallest compass" [Hitchcock 2014].[1] 1

This article introduces *Seshat: Global History Databank* as just such a macroscope and thus as a modern and global variant of our imagination of Seshat's scroll.[2] Seshat will be a comprehensive dataset covering human cultural evolution since the Neolithic, with a well-defined methodology for interrogating and analyzing the data. Seshat is thus a very powerful tool to tackle big questions that play out over (very) long time scales whilst allowing users to drill down to the detail and place every single data point both in its historic and historiographical context. Seshat offers its users a rigorous methodology to actually do *longue durée* history. This article argues for the need for humanists and social scientists to engage with such a data driven *longue durée* history. This article puts forward in detail the Seshat methodology for the reader to grapple with, endorse or criticize. It offers a framework for truly interdisciplinary work integrating history, archaeology, social sciences, evolutionary studies and computer sciences. We also argue that Seshat offers a novel way to connect macro and micro levels of analysis in one framework, that Seshat offers new ways to navigate many of the challenges involved and that Seshat, at the very least, contributes to a better understanding of the issues at stake. 2

The article opens with explaining how Seshat fits into the wider intellectual landscape of both the digital humanities and *longue durée* history. Next, it summarizes the main challenges any "macroscope" faces and how the awareness of these challenges heavily influenced the architecture of Seshat. The article then introduces Seshat at different levels of abstraction. First, it presents Seshat as a platform for existing and future research projects. Second, it explains in more detail how the connection between the Seshat methodology and one project focusing on the evolution of social complexity has been made. Finally, three short case studies are presented detailing how Seshat can be used to study a range of research questions at different scales of analysis. The article thus moves from the macro to the micro and as such it mimics the zooming in of the macroscope. 3

Digital humanities and *longue durée* history

The availability of computational power and the increase in ease of use of this power is changing history fundamentally as an academic discipline.[3] This is most evident in the ever larger historical data sets available to historians, the relative ease with which these data sets can be put together, linked and shared, and the increased popularity of relatively new, for humanists, types of analysis and visualization, ranging from distant reading, over statistical and spatial analysis, to mathematical and computational modelling.[4] The rise of the "digital" pushes historians into new areas of research and away from micro histories centering around well understood archives, texts, datasets and linked to relatively short time scales and played out in relatively small places. The digital also leads to a profound change in how historians conduct research. There is a shift away from lone scholars and toward collaborative work where different collaborators bring different skill sets and knowledge to the table ranging from domain expertise and close reading skills, to synthesizing abilities and analytical skills. 4

This profound impact of computational power and the "digital" on historiography is most notable in the emerging revival of the study of change over (very) long time scales. Whereas the exact time scale can differ if historians embrace Big History, Deep History, write the history of the Anthropocene or embrace the temporally less defined *longue durée* history, it is clear that historians have begun to embrace longer time scales [Christian 2011] [Smail 2008] [Smail et al. 2013] [Guldi et al. 2014].[5] Although the plethora of recent calls to arms for studying history over longer time scales has not yet been matched by an equally rich body of literature actually studying change over such time scales, the debate about its value, methodology and the challenges faced has definitely been opened. The expectations of proponents of the *longue durée* are sky high. In their *History Manifesto*, for example, Jo Guldi and David Armitage see engaging with the *longue durée* as the way for historians to regain an authoritative voice in the public debate, as the way to speak truth to power and thus to restore the influence of history as an academic discipline [Guldi et al. 2014, 12–13]. Whereas they passionately put forward this vision, how exactly it can be achieved is much less clear and this lack of specifics contributed greatly to the debate following the publication of their volume [Cohen et al. 2015] [Hunt 2015] [Lamouroux 2015] [Lemercier 2015] [Moatti 2015] [Trivellato 2015]. 5

As long as there are no clear methodologies to offer on how to do *longue durée* history the debate between proponents and sceptics of *longue durée* history will not move forward. As long as there are no standards on how to put together structured time series which work equally well for, let's say, the Neolithic, the Axial Age or the modern period few historians will actively engage with *longue durée* history. As long as there is no framework which connects macro and micro levels of analysis historians will be worried that the *longue durée* is in part inspired by a (social) science envy and that the tool kit of advanced statistical models is not as well suited to discovering meaning as well as the humanists' skill sets centered around close reading and reading against the grain, honed over decades of practice [Hitchcock 2014]. Will there be a place for the small in the rush for the big, the long, or the global? [Bell 2013] [Bell 2014]. This scepticism is also coupled with a "turn" fatigue. Has the *longue durée*, which cannot be but heavily intertwined with the "digital" and "global" turns, more staying power than the previously equally hyped "cultural", "linguistic" and "spatial" turns? This healthy criticism has greatly informed the methodological set-up of Seshat. More specifically, the methodology of Seshat was defined to address the following challenges: 6

- How do we ensure that a macroscope like Seshat allows analyses at multiple scales?
- How do we make sure that Seshat is not only producing interesting looking visualisations but also allows users to engage with crucial research questions and long standing historiographical debates?
- How can we make sure Seshat not only engages with existing debates but also allows for new questions to be asked?
- How can we ensure that Seshat not only points out connections and influences but also brings real explanatory power to the table?
- How can we assure that Seshat works equally well for understanding the gaps in our knowledge and data sets as it works for visualizing our knowledge?

- How can we assure that the much touted interdisciplinarity does not mean in practice a social sciences project with a humanist touch?
- How can we ensure the intellectual staying power of Seshat by linking with the core task of the historian: explaining change over time?
- How can we strike a balance between the need for universal or regional models and place and time specific historical data?
- How can we assure data quality given the massive amount of data and the very large numbers of collaborators?

Philosophy and architecture of Seshat

7 Seshat is neither the only nor the first research infrastructure project focusing on historical Big Data. There are quite a number of impressive infrastructure projects each with their own distinct and often complementary philosophy and methodology, and each addressing different user needs. For example, in terms of the volume of historical data the Europeana project and the Hathi Trust Digital Library bring together unrivalled amounts of historical data. Other projects or platforms welcome individual researchers to contribute their own dataset and provide help with data management requirements (e.g. tDAR - The Digital Archaeological Record) or provide a platform to link different geo-historical datasets (e.g. *Système modulaire de gestion de l'information historique* – *SyMoGIH*). The goal of the very impressive CHIA Collaborative for Historical Information and Analysis and the associated World-Historical Dataverse project is to bring together existing datasets focusing on the past five hundred years (http://www.chia.pitt.edu/; http://www.dataverse.pitt.edu/)..[6] Finally, although founded in 1949, the Human Relation Area Files - HRAF project still stands out because of its ambition to bring together a very substantial collection of ethnographies and it offers users interested in cross-cultural research added value by providing an entry into the data through a well developed subject classification system. Seshat too offers a very distinctive methodology which is detailed below and which is geared towards specific and quite different user needs. To the best of our knowledge only the more recent Vancouver based Database of Religious History project shares broadly similar aims to Seshat but addresses these in different ways and focuses exclusively on the sub-domain of religion.

8 The aim of Seshat is to be the premier port of call for rigorously testing theories with historical and archaeological data. Using Seshat involves a major shift in research practice for historians. Most current historiography aims to offer new explanations by engaging intensively with existing historiography. Borrowing, endorsing, modifying or challenging existing explanations lies at the core of the historical profession. Adding one additional and more convincing explanation based upon a novel connection between theory and data, stemming from the use of new archival material or a fuller mastery of the sources, the application of more sophisticated analogies, or a superior reading against the grain of the sources is seen as the hallmark of the historian. Seshat enables us to turn this logic around. By embracing the scientific method the principal goal is not necessarily to add one additional explanation but rather to weed out existing explanations which fail on empirical grounds. Seshat enables us to point out the limited explanatory power of existing theories and thus to discard them in preference to other theories with greater explanatory power. By thus reducing the number of theories the aim is to reach temporary consensus about one or a few stronger explanations until new data or new types of analysis restart this process of relative consensus building. [7]

9 Seshat stores data for theoretically motivated variables for a number of units of analysis linked to groups.[8] The list of variables and the hypotheses and predictions these variables allow to test are put together during the initial stages of a research project and are thus declared before the start of data analysis.[9] The units of analysis include polities, quasi-polities, sub-polities, *NGAs* (or Natural Geographic Areas of roughly 100 by 100 kilometers with a relative uniform environment), cities, religious traditions and, allowing for a high degree of flexibility, interest groups such as bands of warriors, trading companies, or religious cults [Turchin et al. 2015, 91–94]. It is possible to add further to this initial list of units of analysis. All data in Seshat is query-able both temporally and spatially and it is therefore possible to bring together data linked to the different units of analysis. Temporally the smallest time step the data can be queried for is a year. Spatially all data will be linked to a GIS shapefile taking the form of a point, a line or a free form polygon (e.g. the border of a polity or the border of an area in which a certain ritual is performed). This allows, for example, for pulling up a data sheet for hundreds of variables for each individual year of the lifespan of a polity whilst linking this data to the geographic extent of the polity for that year. Thus, when using Seshat fully zoomed in data can be viewed at a granularity of one year for a single geographic location. When using Seshat fully zoomed out its temporal scope encompasses the neolithic to the present and its geographic scope is global. The main challenge when using Seshat to analyze data over long periods of time or from large geographic areas is to make sure that the variables work equally well for units of analysis from very different time periods or geographic areas. This is achieved

through an extensive feedback process between variables and data early on in the lifetime of a project.[10] Another key challenge is to ensure that Seshat works equally well for data-rich as for data-poor periods and areas. Seshat has been configured in a number of ways to ensure that data from data-poor periods and areas can still be queried successfully and that a comparison with data-rich periods or areas is meaningful. Firstly, the Seshat data model treats the presence of a certain trait or feature, its absence, and even the "unknownness" whether a trait or feature was present or absent as equally valuable information which can all be used for statistical analysis. Secondly, Seshat employs the approach in which a certain fundamental variable is "proxied" by a number of other, more easily observable measures. A certain degree of redundancy resulting from this approach is a design feature. Even though some proxies cannot be coded, due to lack of information, typically other proxies for the same underlying variable may be available. This information can be used in a statistical analysis to make inferences about the underlying variable even under situations of many missing data. As long as there is some data on some proxies it is therefore still possible to put together a time series reflecting the dynamics of the fundamental variable we wish to study. Modern methods of statistical analysis, such as multiple imputation [Rubin 1987], allow us to make valid inferences about the dynamics of historical variables, as long as at least some of the proxies are known.

10 Each variable generates three types of data. First, each variable has a machine readable bit of code which is most typically a number or an "absent/present/unknown/no data" code. This machine readable code can reflect levels of uncertainty about the data (e.g. domain experts do not know the exact date when a certain feature emerged but agree that it emerged in a certain date range) and levels of disagreement among scholars on the data and its interpretation. Second, each variable is tied to a short (often a paragraph or two) descriptive text which explains the code, qualifies the levels of uncertainty and disagreement, and provides the reader with the necessary contextual information and historiographical background. Taken together these descriptive texts can be read very much like an encyclopaedia entry in which a polity or archaeological sub tradition is introduced to the reader in a structured format. The texts thus summarize in a structured way often very large historiographies which allows the reader to appreciate the depth of knowledge on polities, archaeological sub-traditions, *NGAs*, or, conversely, understand better the gaps in knowledge. Finally, the data is tied via a series of footnotes to the literature (scholarly books and articles, generally the secondary and tertiary literature) which allows the reader to contextualize every single data point even further by providing links to the most recent historiography. As the data will be periodically augmented to reflect new insights the descriptive texts and footnotes will also reflect the evolution in our collective understanding of any of Seshat's units of analysis.

11 Data is uploaded into Seshat in three different ways. Data can be uploaded directly by domain experts like historians, archaeologists or religious studies scholars. Ultimately each data point will receive input from more than one domain expert. Reflecting the endorsements, challenges or debate among domain experts is a crucial aspect of the descriptive text. Experts are also used to lend authority to "obvious" codes on which there is no scholarly debate but for which it is impossible to find a reference (e.g. we code the prevalence of gun powder in the section on military technology variables as "absent" for the neolithic period). Secondly "low-hanging fruit" data which can be identified through a short engagement with the literature is uploaded by research assistants. This data is then presented to one or more domain experts to be approved, augmented, qualified or rejected. Thirdly, an impressive range of digital tools are being developed to help both experts and research assistants to populate data fields more quickly.[11] These tools, especially web scrapers which query very large data collections like Jstor or Google Books for likely candidates - i.e. paragraphs or pages which will most likely contain relevant information on a certain variable – will be an integral part of Seshat. By integrating these tools within the Seshat environment experts and research assistants can cut speedily through very large historiographies. By keeping track of which candidates were helpful and which candidates were mere noise for each variable the algorithms underpinning the process of candidate generation will improve progressively over time.

12 Although at heart a digital humanities and social sciences project Seshat deliberately embraces computational power and thus computer science at all stages of the research process. This sustained engagement with computer science is not done through a piecemeal approach of adopting a specific technology for every step of the research process but by underpinning Seshat with the Dacura data curation platform.[12] Dacura provides a platform upon which to build an integrated Seshat environment for data gathering, data storage, data querying and exporting, and data analysis and visualization. The database itself is a triple store using Linked Data/RDF technology. The digital tools used to facilitate the data gathering process will be integrated into the Seshat environment. Specific work environments are set up for the different types of Seshat users, including work environments for the editor, the domain-experts, the research assistants, and the volunteers. All metadata reflecting the research process (e.g. who uploaded what data when, who augmented, challenged the data, which tools were used) is captured, analyzed and used to improve the data gathering process. For example, it will be possible to assess the quality of the data gathered, on a per-variable basis, with the help of the digital tools. Getting a grip on these metrics is essential to fine

tune the processes or algorithms that produce low quality data. Seshat data is accessible through a number of different outputs including table format and a browse-able wiki-style text-based web page version of the dataset. Whereas the table format invites the users to statistically analyze and visualize the data, the text version opens up the data for a reading and browsing centered exploration and thus allows for a more serendipitous research process. Finally, as Seshat is using Linked Data technology, the data can be linked to other databases with either different units of analysis or using a different granularity.[13] As a result, Seshat data can be easily linked to external data sets and also functions as an entry point into other historical data sets.

The evolution of social complexity

13 Seshat is currently used to test a range of research questions, and associated hypotheses and predictions. Current projects using Seshat focus on the evolution of social complexity, on the possible historical deep roots of areas that today experience the greatest economic growth and political stability, and on the role axial age religions play in explaining social inequality.[14] In this section, we focus on the project studying the evolution of social complexity as this project is furthest advanced and therefore well suited to illustrate how to move from an overall research question to a specific data gathering strategy.

14 Over the past decades a range of explanations of *human ultrasociality* - our ability to cooperate with genetically unrelated individuals - have been proposed [Turchin et al. 2012]. Many of the explanations favour the resource base as the main driver behind the evolution of social complexity. Others focus on the role of warfare and others still stress the role rituals play in producing social cohesion. Systematic empirical research has been lacking to test the strength of these explanations. By using Seshat we want to compare the explanatory power of each of these theories head-on by testing them with systematically collated historical and archaeological data. For this purpose over 600 theoretically motivated variables focusing on social complexity, resources, warfare and ritual have been selected. Data is collected for each of the variables for each year for as far back in time as possible. For the Konya Plain for example this means that data is collected for over 9000 years. For other areas, like Iceland, the time series is much shorter. As the greatest explanatory power for this type of research question lies in studying change over time this particular project did not compromise on the temporal dimension and it aimed at collecting as long a time series as possible. Depending on the specific research question it is possible to limit the temporal coverage. In order to achieve a satisfactory geographic coverage within the lifespan of the grant supporting this project, we have focused on collecting data for a widely spread sample, based around 30 Natural Geographic Areas, which cover more than 400 historical polities. For each of these NGAs data has been gathered for the full lifespan of every polity or archaeological sub culture which was present in or ruled over the NGA. Two parameters shaped the selection of the sample of 30 Natural Geographic Areas. Firstly, the globe was divided into 10 large world zones (e.g. North America or the Indian subcontinent) and for each zone 3 NGAs have been selected. Secondly, these three NGAs for every world zone contain one NGA where social complexity rose very early, one NGA with a very shallow history of social complexity, and one NGA with an intermediate history of social complexity. As a result the world sample of 30 contains NGAs with a very long history of social complexity like Upper Egypt, Mesopotamia and Middle Yellow River Valley and NGAs with a much more shallow history of state formation like Iceland. For NGAs with a long tradition of social complexity this means that we are coding on average 30 polities. For NGAs where social complexity is a relative recent arrival this can yield as little as four or five polities.[15] Taken together these thirty time series represent a very large data set which allows us to test the selected hypotheses and predictions in a statistically rigorous way.

15 Whereas the paragraphs above detail how a project can make full use of the Seshat infrastructure to collect data, the focus of the following sections shifts from the "how" to the "why" and provides three examples of questions that can be better understood through Seshat. Especially the relevance of having access to large structured datasets that can be queried on different timescales and how this allows the scholar to engage differently and with confidence with some of the big historiographical debates will be highlighted. Together these three case studies showcase the relevance of Seshat for historiographical debates playing out at different time and geographical scales.

Testing Theories about the Evolution of Economic Growth and Political Stability

16 The first example of such a big debate - and a debate taking into account long timescales globally – focuses on the historic roots of the staggering degree of inequality in economic performance and effectiveness of governance among nations that we see today. Understanding the causes of these disparities is one of the greatest intellectual puzzles in the social sciences and the humanities. While there is broad agreement about the empirical patterns (which countries are rich and stable, and which are poor and/or prone to political instability), causal mechanisms

responsible for these patterns are much in dispute. Economists have traditionally emphasized capital accumulation and technological progress, as well as policies and incentives affecting factor accumulation and innovation [Rostow 1960]. In more recent years, the attention has shifted to the institutional framework [North 1990], with some economists arguing that economic growth and material improvement can only occur by developing inclusive institutions enabling broad sections of the population to participate in economic and political activities [Acemonglu et al. 2002]. The historical development of particular institutions can have far-reaching consequences [North et al. 1989].

17 On the other hand, some argue that there is a direct effect of geography on economic growth, focusing on such mechanisms as disease burdens [Sachs et al. 2002]. Others have highlighted an indirect effect of biogeographic conditions on current wealth, mediated by the timing of the agricultural revolution in different regions [Diamond 1999]. More recently, economists have gained new insights on these issues by focusing on the ancestral composition of current populations. In particular, Spolaore and Wacziarg have emphasized the historical roots of long-term barriers to the diffusion of innovations in modern times [Spolaore et al. 2013]. What many of these approaches have in common is that they are looking to the past to explain today's inequality in economic performance and effectiveness of governance.

18 Modern evolutionary theory provides a new way of thinking about economic issues [Bowles 2004] [Beihocker 2007]. Explaining change over time (and thus historical data) lie at the heart of this approach. Modern evolutionary theory can also act as a unifying framework allowing us to synthesize different perspectives on political and economic development and design empirical tests of competing hypotheses. Political and economic development are intimately linked and can be seen as two different manifestations of the same deep structural process—building viable states and vibrant economies requires huge numbers of people to cooperate on a very large scale [Seabright 2004]. The ability of humans to cooperate in huge groups of genetically unrelated individuals, or *ultrasociality* [Campbell 1983] [Richerson et al. 1998] is on a scale not seen elsewhere in nature. A key aspect has been the cultural evolution of norms and institutions [Bowles 2004] [Fukuyama 2011] [Richerson et al. 2012] which are characterized by a tension between the benefits they yield at the higher level of social organization and the costs borne by lower-level units [Turchin 2013]. However, this insight raises further important questions: How do these institutions develop over time? Do political institutions precede or follow changes in economic institutions? Ultimately what ecological and historical factors favour the evolution of ultrasocial institutions? To what extent are there universal features of successful political and economic systems? How do these institutions spread?

19 We have many theories purporting to explain economic growth and political stability, thus, but there is little consensus about which of them are correct (or which combination of postulated mechanisms provides the best explanation of the observed patterns). Adjudicating between competing theories of economic performance in a more rigorous and systematic way is critical and this is where the value of Seshat lies. Thus far theories of long-term political and economic development have not been thoroughly tested because we lack data of suitable quantity and quality. Two main problems prevented adopting a Seshat approach in the past. The first problem that plagued previous empirical analyses is the use of modern states as geographical units, even though they may have little relation to historically appropriate units of analysis. For example, one of the spatial units in the database of historical GDP estimates, constructed by Angus Maddison [Maddison 2007], is the USSR. This territory is a very inconvenient unit of analysis for any historical period before the Russian Empire emerged as a Great Power during the eighteenth century.

20 The second problem is how to deal with time. Existing databases suffer from a variety of limitations. Some of the best ethnographic resources, (e.g. the Standard Cross-Cultural Sample, eHRAF World Cultures) are cross-sectional and lack time-depth, while good data on institutions is generally only available for modern societies (e.g. the Political Risk Services database looks back only to the 1980s). Yet to understand causal mechanisms of long-term persistence and reverses we need systematic, long-term dynamic data that tell us how different aspects of societies change with time. Resources that try to overcome these problems are limited to archaeological data (e.g. eHRAF Archaeology), or are constructed by social scientists interested in testing particular theories (e.g., Polity IV). These data yield valuable insights but the number of cases or the time span that can be coded is limited. Additionally, these social scientists are not experts on the societies they code, and thus their databases do not reflect the best current knowledge of expert historians. A similar limitation applies to the full-text HRAF collections, which require researchers to code variables themselves. Some intrepid social scientists have attempted to include deep history in their analyses. For example, Comin and co-authors [Comin et al. 2010] investigated how technological development in 1500 CE (and before that around 1 CE and 1000 BCE) affected the wealth of modern nations. Thompson and Sakuwa look even deeper in time, to 8000 BC [Thompson et al. 2013]. These authors should be commended for attempting to research the deep historical roots of modern economic growth. However, the jumps between 8000 BC and 1000 BC, or even between 1

CE and 1500 CE are huge. Much interesting history happened during these periods, but practical databases that allow to include these developments into analyses have been lacking in the past.

21 Seshat will allow scholars to contrast how well rival theories predict trajectories of actual historical societies in many different regions of the world (and over many different historical periods). Those theories whose predictions are less supported by the data compared to the predictions associated with rival theories will lose their appeal and as empirical evidence against them accumulates, will be rejected. Our understanding of today's inequality in economic performance and effectiveness of governance will thus be on a much firmer footing using Seshat.

Locating Egyptian history in a global historical context

22 The second debate – how to place Egypt in the historical context of the wider region (and indeed globally) – has a more specific timeframe and geographical scope. Egypt has always had an unusual position within the field of History. Broad cultural description has dominated the literature, rather than, for example, an institutional analysis of Egyptian agriculture or the building of a model of the pharaonic economy.[16] As a result of Egyptology's general (there are important exceptions) reluctance to engage more broadly, the big question in macro-historical terms – whether we can characterize Egyptian civilization as African, Mediterranean, or Near Eastern – is still open for debate. Cultural approaches have left a good deal of historical change over time unaccounted for. Traditionally, for example, the entire first millennium BCE, a time when outside groups politically dominated Egypt, has been neglected. That has begun to change thankfully. To be sure Egypt was isolated from the Mediterranean by its main environmental engine, the annual flood of the Nile driven by the monsoonal rainfall in East Africa which was subject to considerable variability for several climatological reasons [Hassan 2007]. The Nile flood, not the king, was the real despot in Egypt. But the interactions between Egypt, the Near East and the Mediterranean world were both complex and important and in many cases were drivers of important change within Egyptian civilization (e.g. The Hyksos domination of Egypt in the 17th and 16th centuries BCE, and Greek dominance from the late 4th to the end of the 1st centuries BCE, to name two obvious cases).

23 In order to start answering the question of Egypt's place within the wider region and in the context of global history we need to access the existing historical data in a format very different from that which drives much of the scholarship: the organization of the chronology by ruling families, or dynasties established by the Egyptian priest Manetho in the third century BCE. The Seshat project, by aggregating all knowable data across a very wide range of variables, allows us precisely to question the conventional chronology and ask new questions. Can the dynastic chronology be replaced by a new convincing data driven chronology? Can we introduce a more dynamic chronology by merging different timescales? As Seshat brings the work of very different specialists together into one framework it is possible to analyze causal inferences between very different variables and place Egypt within a world civilization framework from which it will be easy to compare civilizations over time and space. For the first time we will be able to see how Egypt compares to the classical world synchronically, or to other places, and at the same scale. This will allow historians to ask very important questions in comparative history. For example, were Egyptians of the fourth century BCE worse off than their counterparts in Athens? This is an important, perhaps one of THE important questions in comparative history of the ancient world - what was the role of institutions in economic performance? Were people better off in one political system versus another? This question, for the ancient world, has received renewed attention in recent years. Almost all of this attention has been focused on the classical world. The argument comes down to an institutional one: did Greek society, usually with a heavy emphasis on Athens, create a wider distribution of wealth, i.e. was the distribution of wealth more equal than in societies governed by non-democratic institutions? Did this in in turn lead to greater sustained real economic growth? [Ober 2015] Did religion or bureaucracy act as breaks to social development? And what about abrupt climate change and disease?

24 As the potential riches for understanding Egyptian history offered by the Seshat approach are evident, this begs the question why this has not been attempted before. What are the factors that prohibited the adoption of such an approach in the past? The most important reason is that, as in other areas of humanistic research, Egyptian history has been a text-based field, and the texts, covering four stages of the language and three different scripts from the origins of Egyptian civilization to the Roman conquest, are often difficult. This leads to specialization in particular historical periods or phases of Egyptian history. The assessment of change over time is more often left to archaeologists, whose work is often separated from the language-based studies of traditional historians. Often still, very good archaeological results have not been integrated into historical narratives about Egypt. Needless to say understanding change over time can be obscured, or simply missed, by such text-based approaches.

25 A very large amount of data has currently been entered into the Seshat database for the full sweep of polities that were present or ruled Egypt from the Neolithic until the Ottoman Empire. Seshat, by its very nature, thus clarifies what

we know, what we do not know and what is still debated among scholars, and it highlights the range of evidence that is problematic, disputed and so on. Seeing the total picture of what we do and do not know about these polities, highlighted by comparing the data with data from other places, forces one to think in novel ways about causality. The preliminary results in analyzing this data will be valuable in setting new research agendas. As these research agendas connect historians with their core business of explaining change over time rather than updating existing explanations with the latest intellectual fashion the *longue durée* has staying power and is thus different from most other turns. The Seshat project has already revealed the potential for completely rewriting Egyptian history by assembling an unprecedented and vast number of socioeconomic, political and agrarian variables. Being confronted with this data in conjunction with other civilizations, say for example China, essentially the totality of what we do and do not know about ancient Egypt over several millennia, forces the historian to reconsider why and how Egypt matters in a more global context. The result will allow us to map novel connections between, for example, kings, fiscal institutions and society and it will allow us to identify the unique Egyptian solutions to establishing a political equilibrium.

Modes of religiosity theory in historical context

The third case study demonstrates the usefulness of Seshat for testing theories in the social sciences with historical data. Such data often pertain to particular behaviours occurring in regionally and temporarily bounded groups. This case study demonstrates that Seshat is not about aggregating out, and thus reducing, all data to the level of general trends at the overall polity level. This case study thus highlights the vital importance of "small" data, the importance of the specific context of data to understand key processes in history, and thus represents a macroscope well zoomed in. 26

Scholars have long appreciated that participating in collective rituals increases group cohesion but social scientists are still only beginning to understand how and why rituals have these effects. One of the most empirically productive new developments in this area is known as the theory of *modes of religiosity* or simply the *modes theory* [Whitehouse 1995] [Whitehouse 2000] [Whitehouse 2004] [Whitehouse et al. 2014b]. This theory maintains that collective rituals tend to cluster around two poles in terms of their frequency and emotionality: that is, they tend to be either highly routinized and relatively low in affective intensity (high-frequency, low-arousal or "HFLA" rituals) or rarely enacted and emotionally arousing (low-frequency, high-arousal or "LFHA" rituals). 27

HFLA rituals often form the core practices of large organizations, such as world religions or popular movements. Because the creed and its behavioural prescriptions are highly routinized they form part of people's general knowledge of what to believe and how to act in order to be an upstanding group member. Such knowledge is stored in *semantic memory* – a collection of somewhat abstract schemas comprising the group's belief system together with a set of general procedural scripts for its distinctive practices. These scripts and schemas represent actors and believers as generic group members rather than as particular individuals and so motivate a form of group alignment known as *identification* that is essentially depersonalizing. Since the group's identity markers are enshrined in dogma they can be spread quite easily by gifted orators and through the profusion of more or less sacred texts. Belief systems of this kind can harden into orthodoxies that are more or less systematically policed through hierarchical and centralized authority structures. Thus, HFLA rituals are linked to a series of other features, both psychological and social, ranging from group identification, rapid spread, standardization of dogma, sanctions for unauthorized innovation, and large, centralized systems of top-down control. This clustering of features has been labeled the "doctrinal mode of religiosity" [Whitehouse 1995] but it is not restricted to religious groups and has more recently been shown to characterize the formation and spread of many large-scale secular groups as well [Whitehouse et al. 2014b]. 28

LFHA rituals are commonly found in small-scale traditional societies, for example in the form of arduous initiations. Rituals involving traumatic or painful ordeals are thought to foster a visceral feeling of oneness, known as "fusion" with the group [Swann et al. 2012]. In contrast with the depersonalizing effects of identification, fused individuals have an almost heightened sense of personal agency when the group is salient – indeed they describe themselves as being strengthened by the group just as they, in turn, believe that they make the group strong [Swann et al. 2010]. Fusion is prevalent in small face-to-face groups that have gone through dysphoric experiences together – it is associated with the bonds of kinship in family groups [Whitehouse et al. 2014b] but also in military units, sports teams, gangs, and other local groups that have a strong sense of shared fate [Swann et al. 2012]. Fusion is thought to result from the sharing of especially salient life-changing events – the sort of experiences that define who you are as a person and that, when shared with others, seem to break down the boundary between the personal and the social self. Unlike identification, fusion is therefore rooted in episodic rather than semantic memories [Whitehouse et al. 2014b]. The distinct episodes that shape the personal self are felt to have shaped also the group, such that these 29

unique experiences are equally defining for both. When people fuse with a group in this way they experience any kind of threat to the group as a personal attack and this motivates willingness to fight and die to protect group members [Swann et al. 2010]. Early research on the effects of dysphoric ritual on episodic memory and social cohesion tended to focus on the religious rituals of small-scale societies, labeled the *imagistic mode of religiosity* [Whitehouse 1995]. But as with the doctrinal mode, these imagistic practices are found in highly fused secular groups as well, especially those engaged in risky intergroup conflict. It is thought that the reason why dysphoric initiations are so common in warring tribes and modern armies is precisely because they fuse military units together, creating more motivated fighters. Recent research on both conventional forces and insurgent groups supports this view [Whitehouse et al. 2014a].

30 Much research on the modes theory has focused on the proximate mechanisms linking processes of memory formation to patterns of group alignment and social organization. But the theory also raises questions of ultimate causation: What are the functions of LFHR or HFLA rituals? How do they emerge and fade in the history of different groups? Are there selective pressures that favour one mode over the other, or both at once? Such questions have sometimes been prompted by unexpected empirical discoveries. For example, careful analysis of more than 645 rituals from a sample of 74 contemporary cultures worldwide reveals that LFHA rituals become less common as agricultural intensity increases [Atkinson et al. 2011]. This finding prompted the hypothesis that fusion is especially important in simple societies in which local groups compete for scarce resources. By contrast, in large-scale societies producing storable agricultural surpluses cultural evolution would favour more encompassing identity markers associated with HFLA rituals. We have found much evidence to support these predictions by examining changing patterns of ritual frequency and emotional arousal in the transition from foraging to farming across a range of sites in the Middle East [Whitehouse et al. 2010] [Whitehouse et al. 2013]. The creation of Seshat, however, allows us to pose and answer many more questions about the role of ritual in the evolution of social complexity. For example, do HFLA rituals appear before, during, or after the rise of large-scale, centralized polities? Are armies with LFHA rituals more successful than those that lack them? Do religions that combine HFLA and LFHA rituals last longer than those with only HFLA rituals? Indefinitely many questions of this kind will become answerable once a critical mass of data has been uploaded into Seshat.

31 The questions we are interested in here are only answerable statistically, by quantifying patterns of variation across many polities and religions over long periods of time. But that does not mean the data we use can be rough and ready. Actually, we can only answer our questions adequately if the data are sufficiently fine-grained and precise. The data captured on dysphoric rituals, for example, is often tied to very small groups with complex and changing ties to the overall polities. Capturing these relationships is crucial for our analysis. It is of little use to us to know, for example, that a particular polity has LFHA rituals unless we know exactly what groups were performing the rituals, on what scale, with what degree of frequency, and so on. So our strategy of quantifying history depends on the reliability and depth of qualitative historiography at least as much as it depends on the statistical power of the polities and periods sampled.

32 Fully integrating a qualitative and a quantitative dimension into one data set is a massive undertaking that requires a deep collaboration between researchers with very different skill sets on an unprecedented scale. The Seshat project provides for the first time a methodological framework and technical infrastructure to tackle this task in earnest. Despite the progress made and the impressive synergy already created between the historical and the social sciences, further collaboration between large numbers of scholars is essential for the database to fulfill its potential. Furthermore, there are additional challenges facing a project of this scale. These challenges include finding more methodological robust ways to deal with missing data, understanding better issues of intercoder reliability, and devising more sophisticated ways to capture levels of uncertainty of data and levels of disagreement between experts. As these challenges touch upon core views and attitudes towards data and interpretation in the humanities, tackling these challenges successfully will depend on the level of involvement of humanists. This article has argued strongly for the need to bring different skill sets together into a shared infrastructure. The article is above all a call to the wider community of historians and humanists to keep joining this effort, to make Seshat also their home and to help provide a broad research community with a richer and publicly available dataset to explore the past in novel ways. Data-driven *longue durée* history requires many hands.

Acknowledgements

33 This work was supported by a John Templeton Foundation grant to the Evolution Institute, entitled "Axial-Age Religions and the Z-Curve of Human Egalitarianism," a Tricoastal Foundation grant to the Evolution Institute, entitled "The Deep Roots of the Modern World: The Cultural Evolution of Economic Growth and Political Stability," an ESRC Large Grant to the University of Oxford, entitled "Ritual, Community, and Conflict" (REF RES-060-25-0085), and a

grant from the European Union Horizon 2020 research and innovation programme (grant agreement No 644055 [ALIGNED, www.aligned-project.eu]). We gratefully acknowledge the contributions of our team of research assistants, post-doctoral researchers, consultants, and experts. Additionally, we have received invaluable assistance from our collaborators. Please see the Seshat website for a comprehensive list of private donors, partners, experts, and consultants and their respective areas of expertise.

The corresponding author is Pieter François who wrote an initial draft of the article and who was in charge of the overall integrity of the article. The authors would also like to thank Dr. Alexander O'Conner (ADAPT Centre, School of Computing, Dublin City University, Dublin, Ireland) for his input on the section on Linked Open Data.

Notes

[1]Tim Hitchcock's blog also details how the concept of the "macroscope" saw life in science fiction, how it was picked up by computer science and how it was in turn introduced to the digital humanities via the work of Katy Börner ("Plug-and-Play Macroscopes," *Communications of the Association for Computing Machinery* 54, no. 3 (2011): 60-69. Most recently it featured in the title of a major book on digital history [Graham et al. 2015].

[2] For an overview of the Seshat: Global History Databank initiative and the constituent research projects, see: http://seshatdatabank.info/. For an overview of its methodology aimed at social scientists, see: [Turchin et al. 2015].

[3] Computational power has been used in the humanities for several decades and different labels have been used to describe this use, including "Computational Humanities" and especially the increasingly dominant label "Digital Humanities". For an overview of the aims, history, achievements and criticism on the Digital Humanities, see: [Burdick et al. 2012] and [Jones 2013].

[4] There are several important initiatives to bring together different datasets in one framework and provide the user with novel ways to query and engage with these. An excellent example of such an initiative for a specific time and location is the London Lives project that brings together eight archives and fifteen datasets for the period 1690-1800 in one framework: http://www.londonlives.org/.

[5] Of all the terms referring to the study of change over long timescales, the concept of *longue durée* history is arguably the one historians are most familiar with. Fernand Braudel's classic 1958 article in which he presents his view on the contribution of the *longue durée'* on the discipline of history is still a valuable starting point [Braudel 1958]. The *longue durée* has been most famously championed by the Annales School of historical writing and this to the extent that both concepts are heavily intertwined. For a history of the Annales School, and thus the importance of the *longue durée*, see the classic account by Peter Burke [Burke 1990].

[6] For a broad overview of the overall intellectual agenda driving the CHIA Collaborative for Historical Information and Analysis and the World Historical Dataverse, see: [Manning 2013].

[7] A classic example of a debate where it would be useful to reduce the number of theories is the debate regarding the fall of the Roman Empire. For an introduction to the large number of theories, see: [Demandt 1984].

[8] For an introduction to part of the code book and to many of the clusters of variables Seshat stores data for, see the Supporting Material of [Turchin et al. 2015].

[9]For an example of some the theories and hypotheses pertaining to ritual that will be tested see [Whitehouse et al. 2015].

[10] For example, in order to ensure that the variables worked equally well for all periods in the Upper Egypt time series there was an extensive process of feedback between variables and data and between experts and research assistants. The first phase of this process was a two-year long dialogue between Egypt scholars (John Baines, Elizabeth Frood, and Joseph Manning) and Seshat editors and research assistants. In a second phase a workshop was organized in which the gathered data was discussed and assessed by a range of Egypt scholars (John Baines, Juan Carlos Moreno-Garcia, Brendan Haug, Andrey Korotayev, Joseph Manning, and Julia Zinkina).

[11] These tools are developed as part of the "ALIGNED quality-centric, software and data engineering" project (funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644055; aligned-project.eu). These tools focus upon managing the metadata, incorporating existing structured data into Seshat, and creating candidates from unstructured data/text. A full suite of tools is anticipated to be operational by early 2018. All current Seshat data has been gathered either by domain experts uploading the data directly or by research assistants working under close supervision of a series of domain experts.

[12] Dacura is a data curation system developed by the Knowledge and Data Engineering Group at Trinity College Dublin that provides support for collaborative e-Research applications such as dataset collection, maintenance and publication on the web as Linked Data. Further information is available at the Dacura website, http://www.dacura.scss.tcd.ie or for technical details see [Feeney et al. 2014].

[13] For an introduction of the potential of Linked Data technology, see [Bizer et al. 2009] and for deployment guidelines for humanities and social sciences research projects, see [Van Hooland et al. 2014]. For some examples of Arts, Humanities and Social Sciences projects making good use of Linked Data technology, see: the British

Museum collections site and the Netherlands Institute for Sound and Vision's DIVE event-based browsing of historical media collections online at: http://dive.beeldengeluid.nl/

[14] The evolution of social complexity and the role agricultural productivity, warfare and ritual played in this evolution, is studied as part of the "Ritual, Community, and Conflict" research project (ESRC, UK). The analysis of the deep roots of technologically advanced areas is funded through the research project "The Deep Roots of the Modern World: Investigating the Cultural Evolution of Economic Growth and Political Stability" (Tricoastal Foundation). The role axial age religions play in explaining social inequality is at the centre of the research project "Axial-Age Religions and the Z-curve of Human Egalitarianism" (John Templeton Foundation).

[15]For example, for the Upper Egypt NGA Seshat contains data for 39 polities and for the Kachi Plain NGA Seshat has data for 25 polities, whilst for the Big Island Hawaii NGA Seshat only contains data for five polities.

[16] See e.g. the economic analysis of the origins of agriculture in Egypt in [Allen 1997]. For one proposed model of the pharaonic economy, see [Warburton 1997].

Works Cited

Acemonglu et al. 2002 Acemoglu D., Johnson S., and Robinson J.A. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 91 (2002): 1369–401.

Allen 1997 Allen R. "Agriculture and the origins of the state in ancient Egypt," *Explorations in Economic History*, 34 (1997):135-154.

Atkinson et al. 2011 Atkinson Q.D. and Whitehouse H. "The cultural morphospace of ritual form; Examining modes of religiosity cross-culturally," *Evolution and Human Behaviour*, 32, no. 1 (2011): 50-62.

Beihocker 2007 Beinhocker, E.D. *The Origin of Wealth: Evolution, Complexity, and the Radical Remaking of Economics.* Random House, Cambridge (Mass.), (2007).

Bell 2013 Bell, D. "This Is What Happens When Historians Overuse the Idea of the Network," New Republic (25 October 2013) (http://www.newrepublic.com/article/114709/world-connecting-reviewed-historians-overuse-network-metaphor)

Bell 2014 Bell, D. "Questioning the Global Turn; The Case of the French Revolution," *French Historical Studies*, 37, no. 1 (2014): 1-24.

Bizer et al. 2009 Bizer C., Heath T., and Berners-Lee T., "Linked Data — The Story So Far," *International Journal on Semantic Web and Information Systems*, 5, no. 3 (2009): 1–22.

Bowles 2004 Bowles S. *Microeconomics: Behavior, Institutions, and Evolution.* Princeton University Press, Princeton (2004).

Braudel 1958 Braudel F. "Histoire et sciences sociales. La longue durée," *Annales*. ESC, 13, 4 (1958): 725-753.

Burdick et al. 2012 Burdick A., Drucker J., Lunenfeld P., Presner T., and Schnapp J. *Digital_Humanities*. MIT Press, Cambridge (Mass.) (2012)

Burke 1990 Burke P. *The French Historical Revolution: The Annales School, 1929-89*. Stanford University Press, Stanford (1990).

Börner 2011 Börner K. "Plug-and-Play Macroscopes," *Communications of the Association for Computing Machinery*, 54, no. 3 (2011): 60-69.

Campbell 1983 Campbell D.T. "The Two Distinct Routes Beyond Kin Selection to Ultrasociality: Implications for the Humanities and Social Sciences" In D. Bridgeman (ed.), *The Nature of Prosocial Development: Theories and Strategies*. Academic Press, New York (1983): 11-39.

Christian 2011 Christian, D. *Maps of Time. An Introduction to Big History*. University of California Press, Berkeley (2011).

Cohen et al. 2015 Cohen, D. and Mandler, P. "The History Manifesto: A Critique," *American Historical Review*, 120, no. 2 (2015): 527-554. ("AHR Exchange. On The History Manifesto").

Comin et al. 2010 Comin D., Easterly W. and Gong E. "Was the Wealth of Nations Determined in 1000 Bc?" *American Economic Journal: Macroeconomics*. 2 (2010): 65-97.

Demandt 1984 Demandt, A. *Der Fall Roms: Die Auflösung Des Römischen Reiches Im Urteil Der Nachwelt*. C.H. Beck, Munich (1984).

Diamond 1999 Diamond J. "How to Get Rich," *The Edge* (1999). http://www.edge.org/conversation/how-to-get-rich

Feeney et al. 2014 Feeney K.C., O'Sullivan D., Tai W., and Brennan R., "Improving curated web-data quality with structured harvesting and assessment," *International Journal on Semantic Web and Information Systems*, 10, no. 2 (April 2014): 35-62.

Fukuyama 2011 Fukuyama F. *The Origins of Political Order: From Prehuman Times to the French Revolution*. Profile Books, New York (2011).

Graham et al. 2015 Graham S., Milligan I. and Weingart S. *Exploring Big Historical Data: The Historian's Macroscope*. Imperial College Press, London (2015).

Guldi et al. 2014 Guldi, J. and Armitage, D. *The History Manifesto*. Cambridge University Press, Cambridge (2014).

Hassan 2007 Hassan F. "Extreme Nile Floods and Famines in Medieval Egypt (AD 930-1500) and their Climatic Implications," *Quaternary International*, 173-174 (2007): 101-112.

Hitchcock 2014 Hitchcock, T. "Big Data, Small Data and Meaning," Historyonics blogposts – 9 November 2014. (http://historyonics.blogspot.co.uk/2014/11/big-data-small-data-and-meaning_9.html).

Hunt 2015 Hunt L. "*Faut-il réinitialiser l'histoire?*", *Annales*. Histoire, Sciences Sociales, 70, 2 (2015): 319-325.

Jones 2013 Jones S.E. *The Emergence of the Digital Humanities*. Routledge, London (2013).

Lamouroux 2015 Lamouroux C. "*Longue durée et profondeurs chronologiques,*" *Annales*. Histoire, Sciences Sociales, 70, 2 (2015): 359-365.

Lemercier 2015 Lemercier C. "*Une histoire sans sciences sociales?*", *Annales*. Histoire, Sciences Sociales, 70, 2 (2015): 345-357.

Maddison 2007 Maddison A. *Contours of the World Economy 1-2030 Ad*. Oxford University Press, Oxford (2007).

Manning 2013 Manning P. *Big Data in History*. Palgrave Macmillan, Basingstoke (2013).

Moatti 2015 Moatti C. "*L'e-story ou le nouveau mythe hollywoodien,*" *Annales*. Histoire, Sciences Sociales, 70, 2 (2015): 327-332.

North 1990 North D.C. *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge (1990).

North et al. 1989 North D.C. and Weingast B. "Constitutions and Commitment: The Evolution of Public Choice in Seventeenth-Century England," *Journal of Economic History*, 49 (1989): 803-32.

Ober 2015 Ober J. *The Rise and Fall of Classical Greece*. Princeton University Press, Princeton (2015).

Richerson et al. 1998 Richerson P.J. and Boyd R. "The Evolution of Human Ultrasociality," Ethnic Conflict and Indoctrination. In I. Eibl-Eibesfeldt and F. K. Salter (eds.), *Ethnic conflict and indoctrination: altruism and identity in evolutionary perspective*. Berghahn, Oxford (1998): 71-95.

Richerson et al. 2012 Richerson P.J. and Henrich J. "Tribal Social Instincts and the Cultural Evolution of Institutions to Solve Collective Action Problems," *Cliodynamics: The Journal of Quantitative History and Cultural Evolution*, 3 (2012): 38-80.

Rostow 1960 Rostow W.W. *The Stages of Economic Growth: A Non-Communist Manifesto*. Cambridge University Press, London (1960).

Rubin 1987 Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, Wiley.

Sachs et al. 2002 Sachs J. and Malaney P. "The Economic and Social Burden of Malaria," *Nature*, 415 (2002): 680-85.

Seabright 2004 Seabright P. *The Company of Strangers*. Princeton University Press, Princeton (2004).

Smail 2008 Smail, D. *Deep History and the Brain*. University of California Press, Berkeley (2008).

Smail et al. 2013 Smail, D. and Schryock A. "History and the 'Pre'," *American Historical Review*, 118, no. 3 (2013): 709-733.

Spolaore et al. 2013 Spolaore E. and Wacziarg R. "How Deep Are the Roots of Economic Development?", *Journal of Economic Literature*, 51, no. 2 (2013): 325-369.

Swann et al. 2010 Swann W.B., Gómez A., Dovido J., Hart S. and Jetten J. "Dying and Killing for One's Group: Identity Fusion Moderates Responses to Intergroup Versions of the Trolley Problem," *Psychological Science*, 21, no. 8 (2010): 1176-1183.

Swann et al. 2012 Swann W.B., Jensen J., Gómez A., Whitehouse H. and Bastian B. "When Group Membership Gets Personal: A theory of identity fusion," *Psychological Review*, 119, no. 3 (2012): 441-456.

Thompson et al. 2013 Thompson W.R. and Sakuwa K. "Was Wealth Really Determined in 1800 Ce, 0 Ce, or Even 1500 Ce? Another, Simpler Look," *Cliodynamics: The Journal of Quantitative History and Cultural Evolution*, 4 (2013): 2-29.

Trivellato 2015 Trivellato F. "*Un nouveau combat pour l'histoire au XXIe siècle,*" *Annales*. Histoire, Sciences Sociales, 70, 2 (2015): 333-343.

Turchin 2013 Turchin P. "The Puzzle of Human Ultrasociality: How Did Large-Scale Complex Societies Evolve?," In. Peter J. Richerson and Morten H. Christiansen (eds.), *Cultural Evolution: Society, Technology, Language, and Religion*, MIT Press, Cambridge (Mass.) (2013): 61-73. (Strüngmann Forum Report).

Turchin et al. 2012 Turchin P., Whitehouse H., François P., Slingerland E., Collard M. "A Historical Database of Sociocultural Evolution," *Cliodynamics: The Journal of Quantative History and Cultural Evolution*, 3, no. 2 (2012): 271-293.

Turchin et al. 2015 Turchin P., Brennan R., Currie T.E., Feeney K.C., François P., Hoyer D., et al., "Seshat: The Global History Databank," *Cliodynamics: The Journal of Quantative History and Cultural Evolution*, 6, no. 1 (2015): 77-107.

Van Hooland et al. 2014 van Hooland S. and Verborgh R. *Linked Data for Libraries, Archives and Museums*. Facet Publishing, London (2014).

Warburton 1997 Warburton D. *State and Economy in ancient Egypt. Fiscal Vocabulary of the New Kingdom*. University Press Fribourgh, Fribourgh (1997).

Whitehouse 1995 Whitehouse H. *Inside the Cult: religious innovation and transmission in Papua New Guinea*. Clarendon Press, Oxford (1995).

Whitehouse 2000 Whitehouse H. *Arguments and Icons: divergent modes of religiosity*. Oxford University Press, Oxford (2000).

Whitehouse 2004 Whitehouse H. *Modes of Religiosity: a cognitive theory of religious transmission*. AltaMira Press, Walnut Creek (2004).

Whitehouse et al. 2010 Whitehouse H. and Hodder I. "Modes of Religiosity at Çatalhöyük," In Ian Hodder (ed.) *Religion in the Emergence of Civilization: Çatalhöyük as a case study.* Cambridge University Press, Cambridge (2010): 122-145.

Whitehouse et al. 2013 Whitehouse H., Mazzucato C., Hodder I. and Atkinson Q.D. "Modes of religiosity and the evolution of social complexity at Çatalhöyük," In Ian Hodder (ed.) *Religion at Work in a Neolithic Society: Vital Matters*. Cambridge University Press, Cambridge (2013): 134-155.

Whitehouse et al. 2014a Whitehouse H., McQuinn B., Buhrmester M. and Swann W.B. "Brothers in Arms: Warriors bond like Family," *Proceedings of the National Academy of Sciences*, 111, no. 50 (2014): 17783-17785. Early Edition www.pnas.org/cgi/doi/10.1073/pnas.1416284111

Whitehouse et al. 2014b Whitehouse H. and Lanman J.A. "The Ties that Bind Us: Ritual, fusion, and identification," *Current Anthropology*, 55, no. 6 (2014): 674-695.

Whitehouse et al. 2015 Whitehouse H., Francois P. and Turchin P. "The Role of Ritual in the Evolution of Social Complexity: Five predictions and a drum roll" *Social Evolution Forum*. October 2015 (https://evolution-institute.org/focus-article/the-role-of-ritual-in-the-evolution-of-social-complexity-five-predictions-and-a-drum-roll/).