Andrea Nanetti and Siew Ann Cheong, *Computational History: From Big Data to Big Simulations*, in Shu-Heng Chen (Ed.), *Big Data in Computational Social Science and Humanities*, Springer Series on "Computational Social Sciences". Invited for submission by the Editor in October 2015. Submitted for publication in January 2017. Accepted for publication by the Editor in September 2017. The book is expected to appear in 2018.

Chapter 19. Computational History: From Big Data to Big Simulations

Andrea Nanetti* and Siew Ann Cheong**

Abstract The first section of this chapter gives an overview on how big data and their mathematical calculation enter in the historical discourse. It introduces the two main issues that prevent 'big' results from emerging so far. Firstly, the input is problematic because historical records cannot be easily and comprehensively decomposed into unambiguous fields, except for the population and taxation ones, which are rare and scattered throughout space and time till the nineteenth century. Secondly, even if we run machine-learning tools on properly structured data, big results cannot emerge until we built formal models, with explanatory and predictive powers. The second section of the chapter presents a complex network, data-driven approach to mining historical sources and supporting the perennial historical chase for truth. In the time-integrated network obtained by overlaying all records from the historians' databases, the nodes are actors, while the links are actions. The third section explains how this tool allows historians to deal with historical data issues (e.g., source criticism, facts validation, trade-conflict-diplomacy relationships, etc.), and take advantage of automatic extraction of key narratives to formulate and test their hypotheses on the courses of history in other actions or in additional data sets. The conclusions describe the vision of how this narrative-driven analysis of historical big data can lead to the development of multiscale agent-based models and simulations to generate ensembles of counterfactual histories that would deepen our understanding of why our actual history developed the way it did and how to treasure these human experiences.

* A. Nanetti
School of Art, Design and Media,
Nanyang Technological University,
81 Nanyang Drive, #04-14
Republic of Singapore 637458
e-mail: andrea.nanetti@ntu.edu.sg

** S.A. Cheong
School of Physical and Mathematical Sciences,
Nanyang Technological University,
21 Nanyang Link, #04-14
Republic of Singapore 637371
e-mail: cheongsa@ntu.edu.sg

1 Introduction. The Vision for Computational History

Do historians need computational history to better understand the actual history? Sir Arthur Stanley Eddington (1882-1944), in his 1927 *Gifford Lectures* said that "the contemplation in natural science of a wider domain than the actual leads to a far better understanding of the actual" (1929, 266-267). Before the advent of computational technologies, the value of thought experiments, of which Albert Einstein was very fond, was to present scenarios different from the ones humans observe. The physical scientist would then follow the scenarios through their logical ends to identify what we might have missed and realize what else could be possible if we had lived in a different universe, and ultimately understanding the physical laws that we have at a much deeper level (Eddington 1929).

We believe this can be equally true for simulations in historical sciences. The historical accounts work on "what happened" (i.e., the factual), while computer simulations tell us "what could have happened" (i.e., the counterfactual). Only by combining both the most accurate assessment of what actually happened and what could have happened, we can address the question if in history there are such things as universal laws, from which we cannot deviate in a cause and effect "mechanism-based understanding" (Paolucci and Picascia 2011, 135) of historical phenomena. The power of computer simulations can support historical sciences to develop a shared prescriptive mode of inquiry in the assessment of primary and secondary sources. It will also provide new freedom in the historian's subjective and descriptive identification and assessment of problems to be investigated. Figure 1 illustrates the stages, through which history can improve as a computational discipline.

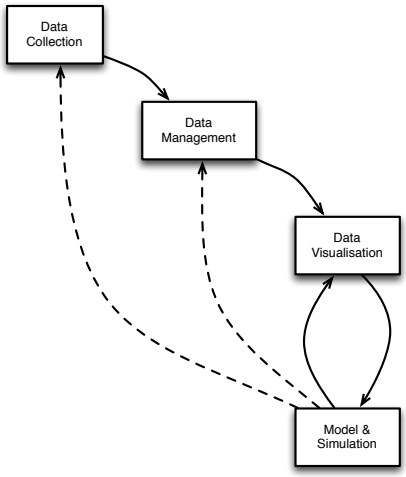

Figure 1. The Stages a discipline must progress through to become computational

In general, it is known that to improve this kind of advancement of learning, historians need to develop specific ontologies to parse data and recognize entities from historical sources. These data can then be mapped into an electronic database and used in analytical environments to build linkages between parsed texts and recognized entities from other heterogeneous sources (e.g., Wikipedia, Open Street Map, etc.) and search engines (e.g., Google Scholar, Microsoft Academic, etc.). For this to happen, historical data also need to be published in online and open access databases, so that they can be properly shared. Historians, as a collective whole, have big digital data, organized in databases but they are not very useful because most of them sit with some kind of organization on the hard disk of individual researchers.

Scholars partially share their data via published books and journal papers, in which data are manipulated in descriptive narratives and need a reverse-engineering process to be used again for a different kind of thinking. Citations and notes are the "procedures intended to communicate an effect of authenticity" (Ginzburg 2012, 21). Since Modern times, historians normally use the footnote as "the one form of proof supplied in support of their assertions" (Grafton 1994, 1995, 1997). However, over time these footnotes can become an unwieldy web that takes considerable effort to navigate. Superhuman efforts are thus required to take all the pieces, and put them together into a recognizable whole.

Therefore, not only the interface to the databases must be properly designed so that it is user friendly, but also and most importantly the data must be curated and tagged by experts using the same identified ontologies and vocabularies, in order to aggregate the data, for example, into a graph database and make it publicly accessible to the international scholarly community, so that any researcher who needs a particular piece of data can find it easily and quickly (e.g., on MSRA Graph Engine, Linx Analytics, etc.). The same identified ontologies and vocabularies can be used to model historical data from historical sources as Linked Data (i.e., best practices to export, sharing and connecting pieces of data in the Semantic Web) and generate, for example, graph representations of the data (e.g., RDF using JSONLD-JavaScript Object Notation for Linked Data), among other solutions (Grinin and Korotayev 2010, Graham et al. 2016).

Unfortunately, nearly all historical databases were designed to be the end products of research projects or programs. To further proceed, databases need to be constantly expanded with the addition of new data sets. Among others, examples of such excellent historical databases include the *Digital Atlas of Roman and Medieval Civilizations* directed by Michael McCormick, the biology-informed *Seshat: Global History Databank* initiated by Peter Turchin, the *Big History Project* conceived by David Christian, *Trismegistos* founded by Mark Depauw, and, to link different databases, *Pelagios* coordinated by Leif Isaksen, and the *Collaborative for Historical Information & Analysis* (CHIA) for creating a world-historical dataset initiated by Patrick Manning with support from the US National Science Foundation (Manning 2013, Manning 2015). We believe that these databases, as well as others, can become portals of historical knowledge, if they also offer functionalities to combine data with metadata, show visualizations of

this combination, and run simulations based on insights gained from such visualizations.

Beyond the mandatory identifying metadata associated with each piece of historical data, databases should also record the interactions between researchers from different disciplines and the data, in the form of metadata. Clearly, these forms of interactions between experts could not happen easily without the computer database, because most of the expert assessments are pre-publication level and conjectural, so we will not see them in journal publications or books, however long we wait. In this sense, having very diverse data made available on a database, and having metadata to augment the data sets themselves is one way the digital computer is revolutionizing the study of history, by allowing historians more intimate interactions with the data, and consequently closer interactions amongst each other.

However, if we stop at this stage, then data sets and metadata will accumulate, and very quickly the volume of data and metadata available will be so large that no one expert can comprehend them anymore. Therefore, to take advantage of the third wave of 'really' computational history's opportunities, historians can be helped by the computer to better comprehend the collection of data and metadata, i.e., to go from simply data management aided by the computer (Graham et al. 2016, 73-111 and) to more sophisticated topic modelling and data visualisations ("deforming, compressing, or otherwise manipulating data in order to see them in new and enlightening ways", Graham et al. 2016, 113-158 and 159-194), and network analysis (Cornwell 2015, Hitzbleck and Hübner 2014, 7-15, and Graham et al. 2016, 195-264).

In this Data Visualization stage, the historian will borrow various machine learning strategies from the computer scientist to discover patterns in the data. Because historians traditionally spend long hours working directly with data, they become very good at formulating hypotheses, and thereafter finding from memory other pieces of data that would support such hypotheses. However, it is highly likely that they miss many other patterns in the data that do not fit into their modes of theorizing. The suite of data visualization and machine learning methods developed by computer scientists over the years can help discover most of these patterns. We feel such methods have been under-utilized because (1) the historical databases are fragmented, and therefore, patterns across different data sets cannot be detected, and (2) the methods are not traditionally included in the training of historians. More importantly, the historical databases are designed for human query, and not necessarily structured for machine query and thus machine learning.

The final stage that history must reach to become a full-fledged computational discipline, is Modelling and Simulation to explore big historical data in big simulations, algorithmically—as John Holland would say (Holland 1975, Mitchell 1996, 2-3). Models can be top-down (equation-based) or bottom-up (rule-based), and can be analysed (by following the chain of logic in the equations or rules until we arrive at conclusions) or simulated (by letting the computer follow the chain of logic, so that we can interpret the conclusions). Models help us understand the big

picture, by functioning (in conjunction with analysis, and/or more likely, simulation, when the model becomes too complex) as a *macroscope* that synthesizes our fragmentary knowledge and insights into a complete whole.

As summarised by Shawn Graham, Ian Milligan, and Scott Weingart, the term *macroscope* was first used by Joël de Rosnay (1979) to discuss complex societies. In literary criticism, a similar concept was called 'distant reading' by Franco Moretti (2005) and 'macroanalysis' by Matthew L. Jockers (2013). As for cultural history, an exemplar demonstration of "data-driven macroscopic" approach is given by Maximilian Schich and his research team (2014, 562). Murray Gell-Mann pointed out in his keynote lecture *A Crude Look at the Whole: A Reflection on Complexity* given at the homonymous international conference hosted by Nanyang Technological University Singapore from 4-6 March 2013, to increase the understanding of historical processes, we should improve the approach pioneered by the British historian Toynbee, rather than simply criticizing and marginalizing. In his twelve-volume magnum opus *A Study of History*, Toynbee presented the development of major world civilizations starting from a history of the Byzantine Empire (Toynbee 1934-1961, Gell-Mann 1997, 9, and Shäfer 2001, 301). Others, like Erez Aiden and Jean-Baptiste Michel (2013) also wrote about "a [macro]scope to study human history" (Graham et al. 2016, 2).

In Section 3 we will explain the limitations of equation-based modelling, however powerful, when applied to historical inquiry, and why it is more natural and appropriate to adopt agent-based modelling (ABM). We will explain how we would go about developing agent-based models, and how we can use their simulations to add to our understanding of history. In spite of it being critical to *macroscope* approaches, ABM, as a computational practice, remains largely unfamiliar to digital historians, despite signs of increasing interest (Gavin 2014, Brughmans and Poblome 2016). In the historical landscape, ABM, like in other disciplines, would explain general trends and offer a complementary, but very different path to macroscopic knowledge. Joe Guildi and David Armitage in their *Historical Manifesto* (2014) argued the importance of macroscopic thinking. Shawn Graham, Ian Milligan, and Scott Weingart gave to their monograph on *Exploring Big Data* (2014) the subtitle *The Historian's Macroscope*.

Ultimately, the purpose of having models is to do predictions, and these can be qualitative or quantitative. If we re-simulate the past, we can end up with a simulated world (Gavin 2014, 24), interwoven by counterfactual histories. If we simulate into the future, we will be exploring different scenarios.

Counterfactualism and the debate over contingency versus inevitability have been explicit themes in modern evolutionary biology since Stephen Jay Gould's book about evolution and how to interpret evidence from the actual past (Gould 1989). The discussion became relevant for history of science, in general (Radick 2005), and Osvaldo Pessoa Jr has been exploring the role for computer models in assessing history of science counterfactuals (Pessoa 2001).

This discussion fits in the discourse of "The Social Logic of the Text", as discussed in 1997 by Gabrielle Spiegel, who argued that "while cultural anthropology and cultural history (together with the New Historicism…) have

successfully reintroduced a (new) historicist consideration of discourse as the product of identifiable cultural and historical formations, they have not been equally successful in restoring history as an active agent in the social construction of meaning" (Speigel 1997, 9). But, before we explain how simulated histories can help historians, let us first link simulations to the historian's key problematics.

1.1 History's Chase for Truth

According to the New Oxford American Dictionary, the Greek word ἱστορία/*historia* comes from *histōr*, which means 'the one who saw, the testimony > learned, wise man', and comes from an Indo-European root shared by wit/vit (to know) that gave Sanskrit *veda* 'wisdom' and Latin *videre* 'see', as well as the Old English *witan* of Germanic origin, and is related to Dutch *weten* and German *wissen* (Joseph and Richard 2003, 163). Thus, history is a kind of knowledge acquired by investigation with the intent to generate wisdom, and implies the action of 'inquiring/examining', which is a requirement to move from knowledge (knowing how to do something) to wisdom (knowing under which situations to act).

If one agrees with Aristotle (Poetics, 51b), the historians speak of that which exists (of truth), the poets of that which could exist (the possible). In a computational modelling perspective, Michael Gavin (2014) notes that "on the surface, computational modelling has many of the trappings of science, but their core simulations seem like elaborate fictions: the epistemological opposite of science or history". He proposes "that these forms of intellectual inquiry can productively coincide" (2014, 1). But, it is not as simple as that. Let's give a few significant examples. Ronald Barthes (1967)—following the structural linguistics of Ferdinand de Saussure and its anthropological extension made by Claude Lévi-Strauss—rephrased this key speculation arguing if "the narrative of past events, subject usually in our culture, from the Greeks onward, to sanction of the historical 'science', [...] is really different, for some specific trait and an indisputable relevance, from the imaginary narration, which we can find in epics, novels, drama?"

On an opposite interpretative angle, we have Carlo Ginzburg. "Under the influence of structuralism, historians oriented themselves towards the identification of structures and of relationships. This identification rejected the perceptions and the intentions of individuals, or turned them into independent experiences, thus separating knowledge from subjective consciousness. In parallel, the number, the series, the quantification, which Carlo Ginzburg has called Galilei's paradigm [1986, 96-125 and 200-213], drove history towards a rigorous formulation of structural relationships, the establishment of whose laws became its mission" (Vendrix 1997, 65). The synopsis provided by the publisher for Carlo Ginzburg's essay collection (2012), states that he "takes a bold stand against naive positivism and allegedly sophisticated neo-scepticism. It looks deeply into

questions raised by decades of post-structuralism: What constitutes historical truth? How do we draw a boundary between truth and fiction? What is the relationship between history and memory? How do we grapple with the historical conventions that inform, in different ways, all written documents?".

Bernard Williams' famous statement that "the legacy of Greece to Western philosophy is Western philosophy" (2006, 3) is particularly true in this circumstance, because Plato's iconic quote from the *Apology of Socrates* (399 BCE) still provides the exact framework: *The unexamined life is not worth living* (Ὁ δε ἀνεξέταστος βίος οὐ βιωτὸς ἀνθρώπῳ, Apology of Socrates, 38a). Life is not worth living without ἔλεγχος/*elenchus*, that is examination, argument of disproof or refutation, dialogue; cross-examining, testing, scrutiny especially for purposes of refutation. Such is the Socratic *elenchus*, often referred to also as *exetasis* or scrutiny and as *basanismus* or assay (Vlastos 1982).

Since Herodotus of Halicarnassus (c. 484–c. 425 BCE) in Classical Antiquity, Lorenzo Valla (c. 1407–1457) in the Renaissance, Leopold von Ranke (1795–1886) in Modern Times, and Marc Bloch (1886-1944) in the twentieth century, the critical assessment of the authenticity and reliability of historical sources is the basic and fundamental tool that historians have been using as a *condicio sine qua non* to acquire their data and establish relations such as cause-effect among them (Galasso 2000, 293-353, Ginzburg 2012, 7-24). While the "procedures used to control and communicate the truth changed over the course of time" (Ginzburg 2012, 231), and the use of the same data can be dramatically different in various accounts bearing on the same past events across time, space, and cultures as well (Grafton and Marchand 1994, Guldi and Armitage 2014, Wang 2016).

Thus, the historians' key problematics have endured for a long period of time. In 1986, Carlo Ginzburg, in his seminal essay on *Clues: Roots of an Evidential Paradigm*, highlighted how history shares with two pseudo sciences, divination and physiognomics, not only roots but also their derivative sciences, law and medicine, that "conducted their analysis of specific cases, which could be reconstructed only through traces, symptoms, and clues. For the future, there was divination in a strict sense; for the past, the present, and the future, there was medical semiotics in its twofold aspect, diagnostic and prognostic; for the past, there was jurisprudence" (Ginzburg 1989, 104-105, and Momigliano 1985).

1.2 The Historians' Big Data in a Computational Perspective

The electronic computer radically changed at all levels the ways our society and economy work (Robertson 1998 and 2003). Historians are fully aware of the importance of this technological turn for the advancement of historical research (Ladurie 1978, Galasso 2000, 311-315, Ginzburg 2001, Cohen and Rosenzweig 2005). In principle, the historian is not refractory to new technologies: all historians went digital, in one way or another. They "have been actively

programming since the 1970s as part of the first two waves of computational history" (Graham et al. 2016, 58).

Today, computers can do for historians what they did, for example, for mathematicians and chemists in the twentieth century, both at the level of capacity of observation and theoretical speculation (Robertson 1998). For example, chemists used to create models of molecules using plastic balls and sticks. Today, the modelling is carried out in computers. In the 1970s, Martin Karplus (Université de Strasbourg, France and Harvard University, Cambridge, MA, USA), Michael Levitt (Stanford University School of Medicine, Stanford, CA, USA), and Arieh Warshel (University of Southern California, Los Angeles, CA, USA) laid the foundation for the powerful programs that are used to understand and predict chemical processes. Computer models mirroring real life have become crucial for most advances made in chemistry today, and on 9 October 2013, the Royal Swedish Academy of Sciences decided to award the Nobel Prize in Chemistry for 2013 to them "for the development of multiscale models for complex chemical systems".

However, after the "Digital Humanities Moment" (Graham et al. 2016, 37-72), when historians started delving into data management and experimenting with various software to shed new light on their data sets, they seem to find it more difficult to take full advantage of the fact that computation itself is again *morphing*, as William Brian Arthur would say (2009, 150-151). Machine learning algorithms, one of computation's key technologies, underwent radical change and have now opened new horizons to the automation and speed of discovery (Domingos 2015). In this third wave of computational history the barriers of entry to powerful computing and big data have never been lower for the historian (Graham et al. 2016, 58). So, it should be more attractive and easier for historians to step in. But, in practice, it is more complicated because the question of the sources—which keeps on being of the essence to the historian's craft at each dramatic technological turn (oral-to-written, handwritten-to-printed, analog-to-electronic, and now from mathematical to algorithmic computation)—is acting as a bottle-neck. Let us explain why and how.

These expanded research capacities can allow new computational-driven research questions (and new answers): What shall the historian do having *all* data available in a digitalized form accessible in any language? What are the implications when *all* research materials are digitized and searchable through metadata in any language? Can we understand the mechanisms of convergence/divergence between local communities and international networks? How can the same networks/people bring new wealth and development, or generate war and poverty? Which dynamics and mechanisms operate in the world systems of individuals, families, cities, and countries? When we know the relationship between *all* (past) facts, *all* their (still present) traces/evidence, and *all* historiographical interpretative accounts, what kind of wisdom can be built on them? Is it possible to model bottom-up universal laws to influence the future? (Nanetti and Cheong 2016, 8).

Since the introduction of punch cards to enter data into computers, historians started to create large data sets that may be analysed computationally. In the 1970s, the French historian Emmanuel Le Roy Ladurie was the first to foresee the implications of the use of the computer in historical studies: "History based on computers/information technology is not limited to a very specific category of research, but also leads to the establishment of an 'archive'. Once transferred to tape or punched cards, and after having been used by a first historian, the data can in fact be stored for future researchers, who want to find non-experimented correlations" (Ladurie 1978, I, 3).

Since then, in their daily research activities, historians are producing and accumulating extremely large digital datasets, in different languages and formats. More and more historical databanks are becoming available on the Internet. Thus, big data are becoming part of the historian's craft, worldwide. As more historical databases come online and overlap in coverage, historians and history as a discipline needs more and more big data approaches to cope with the increasing volume of available sources and interpretations. Despite these big data, so far, big results are at the horizon but not yet clearly visible. Why?

1.3 What Prevented 'Big' Results from Emerging so Far?

Cognitive computing borrows methodologies from two other disciplines, artificial intelligence and signal processing, for the simulation of human thought processes, while computational history aims to simulate the historian's craft, in a computerized model. Being at the very birth of artificial intelligence and automatic signal processing, current scholarship and technology may have science fiction dreams, but cannot have the presumption to automatize history as a whole, because its data volume and complexity are still far beyond any available digital storage system capacity and machine learning capability (Pavlus 2015). Nonetheless, computational history can be extremely relevant to develop a new and more efficient study of primary sources and secondary literature supporting the perennial historical chase for truth.

The bottle neck is the exegesis of the sources, because before dealing with big outputs, we need to work on big inputs. The ontology adopted for the definition of the entities and properties of databases is at the heart of the visualisation processes that can allow agent-based modelling to shed new light on historical records. Thus, computational history, before getting into the debate on the laws and purpose of history (Gilbert 1990 and Popper 1999, 105-115), is called to agree upon standardized methods to define machine-readable ontologies for both data (items known or assumed as facts) and the relationships among data (i.e., information, facts provided or learned about something or someone), which can be automatically extracted from primary and secondary sources, and possibly allow to expand, quantitatively and qualitatively, historical evidence, that is the available

body of facts that the historian uses to judge whether a belief or proposition is true or valid.

In a cognitive computing perspective, this process can be rephrased as provenance-based validation (Wong et al. 2005). In the adoption of such a practice, historical records need to be comprehensively decomposed into unambiguous fields in order to be able to feed machine learning algorithms, which, firstly, can engineer evidence-fact-event relationships in both primary sources and secondary literature, and, secondly, build models of historical phenomena accounts in local, regional, and global historical scenarios (e.g., in our case study, trade-conflict-diplomacy relationships).

Hence, this paper (re)address the question of the sources and aims to provide some solutions and facilitate this new 'macroscopic' computational turn in historical studies. The solution that we propose to fill the gap comes in two stages: (1) to restructure the computation of sources using big data automatic narratives to extract facts from them and see their potential interconnections; and (2) to look at intensity in the flow of facts to identify events as tipping points (Gladwell 2000) in societies' natural nonlinear life using agent-based big simulations.

Firstly, historical data are seen by computer science people as unstructured, that is, historical records cannot be easily decomposed into unambiguous fields, except for the population and taxation ones, which are rare and scattered throughout space and time till the nineteenth century. This fact, in a computational perspective, prevent taxation and population databases to be scalable and aggregated with other datasets. An evident demonstration for taxation records is the *Online Catasto of Florence*. It is a searchable database of tax information for the city of Florence in 1427-1429 (c. 10,000 records uploaded till 1969) based on the work by David Herlihy and Christiane Klapisch-Zuber, Principal Investigators, *Census and Property Survey of Florentine Dominions in the Province of Tuscany, 1427-1480*.

Secondly, machine-learning tools developed for structured data cannot be applied as they are for historical research. Both the exegesis of primary historical sources, and the analysis of how those same primary sources have been selected and interpreted in various historiographical narratives are of the essence in this issue. The historians are required to shift from generalization to conceptualization, because univocal distinctions among theoretical units (e.g., evidence, fact, event) and historical phenomena (e.g., trade, conflict, diplomacy) become necessary conditions to generate new computational ontologies for databases (Guarino et al. 2009) and their application in agent-based modelling for historical simulations (Gavin 2014).

2 Big-Data Automatic Narratives as a prerequisite for Big Simulations

According to Thomas R. Gruber (1993 and 1995), a computational ontology requires a research domain to share an explicit formal specification of the domain terms themselves and their reciprocal relationships. Following Gruber's methodology, Andrea Nanetti extracted from the Morosini Codex (1205-1433) a coherent set of indexing terms (Nanetti 2010, xvii-xix and 1853-2274) to aggregate data for the interactive study of global histories. This research project, started from the world as seen from Venice, is creating an international research team with the ambition to engage the scholars of all other coeval chronicles written in Chinese, Arab, Russian, Persian, etc. (Nanetti and Cheong 2016).

This Venetian beginning is highly relevant in global context for three main reasons. Firstly, the Morosini codex was the model for the subsequent Venetian vernacular historiography leading to the famous 58-volume *Diarii* (1496-1533) by Marin Sanudo the Younger (1879-1902). These primary sources, providing information on all the empires and cities having marketplaces in the inhabited known world (the oecumene), represent one of the most important international texts for late medieval European and Mediterranean history. They deal with innumerable political and economic records taken mainly from merchants' (news)letters and the Venetian council deliberations (Nanetti 2010, xi-xvii). Secondly, the Mediterranean basin has the longest and best-studied record of the ways in which human activities have transformed the world (Abulafia 2011, i-xxxi). Thirdly, in a computational perspective, the time period between 1205 and 1533 provides just enough but not overwhelming data to imagine big simulations (Nanetti and Cheong 2016, 22-25).

The system, to which this interactive study of global histories refers, is the intercontinental Afro Eurasian communication network, which was first investigated in a scholarly and comprehensive way by the German geographer Ferdinand Freiherr von Richthofen (1833-1905) in his magnum opus *China* (1877-1912). In 1876 and 1877, baron von Richthofen anticipated the results of his work in two lectures given in Berlin, at the German Geological Society (Waugh 2007, 3). On 6 May 1876, he significantly chose to dedicate the first one to the sea routes (Richthofen 1876). The second, given in 1877, was about the communications over land (Richthofen 1876).

In this system, the actions (i.e., key relationship among events) have been identified in trade, conflict, and diplomacy. The agents (i.e., the historical actors) chosen for the simulations are in first instance the governments, which allow us to analyse continuity and change patterns in trade-conflict-diplomacy relationships among events at a world scale. On a higher level, this automatic extraction of key narratives from a historical database allows historians to formulate hypotheses on the courses of history, and also allows them to test these hypotheses in other actions or in additional data sets.

2.1 Automatic Source Provenance Identification and Facts-Evidence-Event Validation

As the name implies, the past is an era gone by: it is no longer with us in the present. Historians use traces (the poor remains, still extant in the present) of what happened as clues to select, investigate, and judge events of the past. We think and speak of a past event as *factual* if someone or something we trust provides evidence for it. We consider accounts of such events *truthful* if trustworthy people wrote them down. By the time we read the accounts, we can have a variety of different evidence, from one single record written once in an otherwise proven *truthful* chronicle to chains of endorsements by *trustworthy* people, and therefore we consider such accounts *trustworthy*.

We frequently find two accounts that are highly similar in two or more sources, but with noticeable differences between them. Do these then refer to the same event, or to separate events? For events that appear in some sources but not in others, how would we know they are real? Similarly, for accounts that are highly similar, how would we know if they refer to the same event? Historians learn to judge the authenticity of historical records as part of their training, and become better over time. However, in the era of Big Data, the amount of data and records will overwhelm historians. Therefore, we need the computer to help us validate the historical records if we want the process to be scalable.

To do so, we (re)propose to decompose historical records into their elementary constituents: who, what, when, where, why, and how. All elements must be demonstrably factual before the record can be considered factual. In other words, if the actor reported in an account appears also in other accounts (especially competing ones), the actor is likely to be a real person or a real institution in the past. On the other hand, if the actor appears only in one account, and is imbued with incredible or inconsistent attributes, there is a good chance it is made up. It turns out that checking the consistency in the profile of an actor is non-trivial, because different accounts may refer to the same actor using different names that may sound similar. Similarly, consistency in different accounts can also help us establish the validity of events, locations, motives, and actions.

In the validation process described above, we see that ‘who’, ‘what’, ‘when’, ‘where’, ‘why’, ‘how’ are the basic building blocks of our knowledge about the world. By themselves, they do not amount to much. For example, ‘Marco Polo’ may appear in multiple accounts, and based on this consistency we thus suspect his existence in the past as factual (Orlandini 1913). The consistent accounts thus provide *evidence* for the existence of ‘Marco Polo’. Similarly, ‘Catai’ appears in multiple accounts, in manners that suggest that it refers to a place (Yule and Cordier 1913-1916). We thus establish ‘Marco Polo’ and ‘Catai’ as factual data. This is to be distinguished from non-factual data, which can refer to beliefs, whose contents may not be factual, but their existences are not in doubt.

By themselves, data are not very insightful. As we learn more about the world around us, we start to draw relationships between data. For example, ‘Marco Polo

in Catai' tells us more about 'Marco Polo' and 'Catai', more than the what we can infer from the separate factual existence of 'Marco Polo' and 'Catai' (Orlandini 1926). In the same way, we can understand when relationships are counterfeit. For example, 'Jacob of Ancona'—the supposed author of a book of travels, in which he was assumed to have reached 'Catai' in 1271, four years before Marco Polo—ceased to exist in the historical landscape when his account of 'Catai' was demonstrated to have been forged in the twentieth-century by David Selbourne (Halkin 2001).

We call this level of knowing about the outside world 'information', which allows us to say something about factual data and their relationships. In this classification scheme, historical facts are information decomposable into data entities and their relations. A historical event, though seemingly more complex than historical facts, is a collection of interrelated historical facts, but remains at the level of 'information' that is structured into a narrative. Here let us warn that the 'why' element of a narrative is extremely difficult to validate and establish as fact, because motives frequently depend on the actors interpreting them, while the actor responsible for an action may not provide a written account of its motive, truthful or otherwise. Motives are also notoriously susceptible to reinterpretation in subsequent accounts, for reasons that are difficult to uncover. Establishing the factual status of a motive is thus a major challenge, since the consistency criterion for validation frequently fails.

In the Data, Information, Knowledge, Wisdom (DIKW) hierarchy popularized by Russell Ackoff (1989), knowledge lies above information in our knowing of the world, and wisdom represents the highest level of knowing. In the DIKW hierarchy, knowledge is a collection of information, organized into a procedure, for acting on the world to solve problems. Wisdom is knowing when to act and when not to, because there may be no value in solving some problems, or because we need to prioritize which problem we solve first. The historian's goal for studying history is ultimately wisdom, but to acquire it we must pass through the knowledge stage. To get to this prescriptive and proactive stage of knowing, modeling and simulation is necessary.

But before we describe how to build ABMs based on historical events, and how to simulate these ABMs to obtain counterfactual histories, let us highlight the different ways historians and physical/computer scientists define data, information, facts, evidence, and events. This comparison is shown in Table 1.

	History/Humanities	Physics/Computing
Data	Things known or assumed as facts, making the basis of reasoning (Philosophy).	The quantities, characters, or symbols, on which operations are performed by a computer.
Information	Facts provided or learned about something or someone.	Data as processed, stored, or transmitted by a computer.

Facts	- A piece of information used as evidence or as part of a report or news article. - The truth about events as opposed to interpretation (Law). - The available body of facts or information indicating whether a belief or proposition is true or valid. - Information given personally, drawn from a document, or in the form of material objects, tending or used to establish facts in a legal investigation or admissible as testimony in court.	Synonymous with information.
Evidence	- The available body of facts or information indicating whether a belief or proposition is true or valid. - Information given personally, drawn from a document, or in the form of material objects, tending or used to establish facts in a legal investigation or admissible as testimony in court (Law).	Collection of data demonstrating the reproducibility (consistency) of an information/fact.
Event	A thing that happens, especially one of importance.	A single occurrence of a process (Physics).

Table 1. How historians and other humanities scholars define data, information, evidence, facts, evidence, and events differently from physical and computer scientists.

2.2 Complex Networks Visualisations of Historical Datasets. Trade-Conflict-Diplomacy Relationships as a Key Case Study

Assuming that we have solved the problem of provenance and validation, and have successfully created a database of historical events in narrative format ('who', 'what', 'when', 'where', 'why', 'how'), we now have the problem of extracting insights from such a database. After we have also created an ABM, this would be much easier, because we can use simulations to fill in the gaps and create an animation of the events. However, insights are precisely the ingredients needed to create the ABM, so we are faced with a chicken-and-egg problem. Therefore, in place of the ABM animation, we turn to model-free visualization

strategies to discover the most important stories that can emerge from the database. We do this using a complex-network approach.

First, we create a multigraph where the nodes represent actors (which in our preliminary study are governments), and nodes can be connected by three different types of weighted links, one for conflict, another for diplomacy, and the last for trade. Other actions can also be included in follow-up studies. We start by setting the weights of all links to zero. For a given event, we identify the actors involved, and also the action. For example, in the record of Venice going to war with Rome, the actors are the governments of Venice and Rome, and the action is war. Therefore, we add one to the weight of the conflict link between Venice and Rome. After we have gone through the full list of events in the database, we end up with a time-integrated multigraph (see Figure 2). We can use community detection methods in the complex network literature (Girvan and Newman 2002, Blondel et al. 2008, Alvarez et al. 2015) to identify groups of nodes that are persistently at war, at peace, or trading with one another. If such groups exist, historians must then find cultural and geopolitical reasons to explain them.

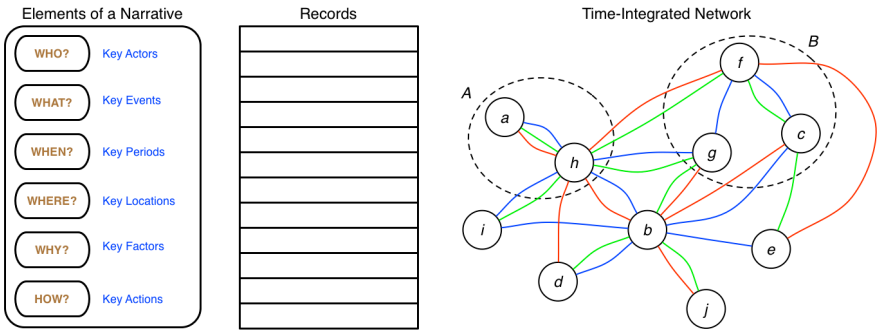

Figure 2. Constructing a time-integrated complex multigraph using the list of records in a historical database. In the database, records are organized into narratives (with elements 'who', 'what', 'when', 'where', 'why', 'how'). In the complex network, nodes represent governments, while links represent actions (red for conflict, blue for diplomacy, and green for trade). The dashed circles represent schematically persistent groups of nodes discovered using community detection methods.

Using this database, it is possible to work on the identification of transient groups of nodes at war, at peace, or trading with one another. To discover them, we need to construct a timeline of complex multigraphs representing different periods in history. Again, patterns that we discover here represent the coarse flow of history, and explanations are called for. Finally, we can perform a *time-resolved analysis* of the database at the event level, to identify the *key events* that form the ingredient for our explanations (Figure 3), and also *key periods* that historians should focus their attentions on (Figure 4).

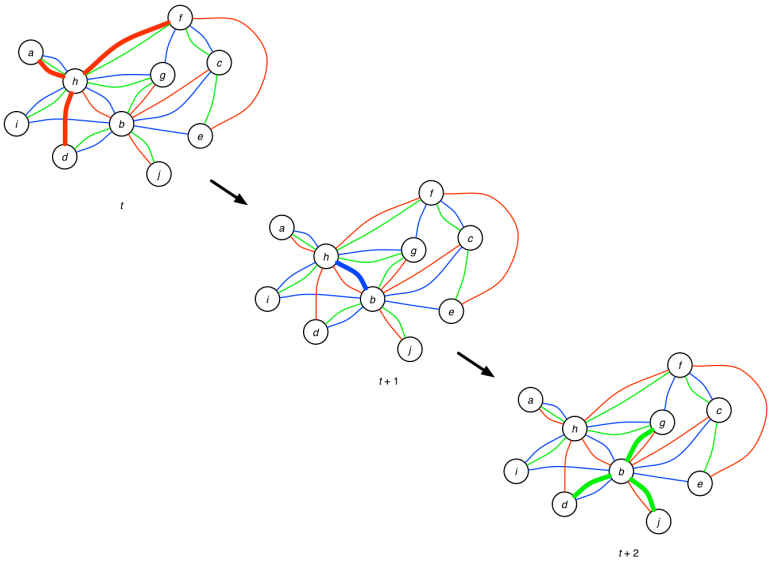

Figure 3. In this figure, we highlight active events during successive time windows t, t+1, t+2 by showing them as thick links. The convergence of events on h in time window t and divergence of events from b in time window t+2 points to the event involving h and b in time window t+1 as a *key event*.

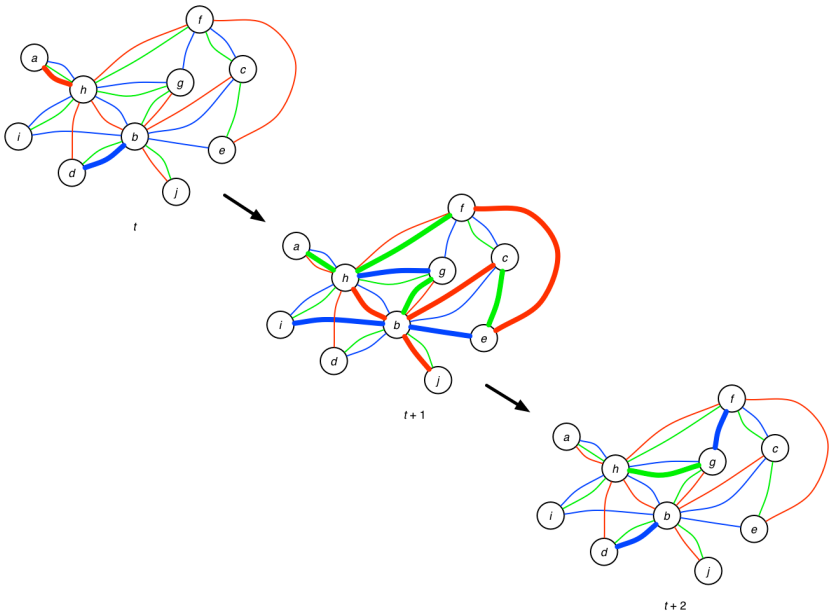

Figure 4. In this figure, we highlight active events during successive time windows t, t+1, t+2 by showing them as thick links. The dramatic increase in number of events in time window t+1 relative to time windows t and t+2 points to time window t+1 as a *key period*.

2.3 Formal and Informal Models. Going from Patterns to Models

From the time-integrated complex network and the time-resolved analysis based on it, we can identify many patterns that we can use to develop large-scale historical narratives. This will feel easy, and the narratives compelling, because we have extracted key narrative elements from the historical database. Without the method of automatic narratives, this historiography would take considerably more effort from the historian. Unfortunately, historical databases today are not designed to support inquiries through such machine learning strategies, however well designed their search tools for enquiries by human historians. To support pattern discovery by automatic narratives and other forms of data visualization, existing historical databases must be systematically reorganized.

However, we must not stop at data visualization and finding patterns, which in some sense represent informal models. In our quest for historical understanding, such patterns are what economists today call stylized facts (Brian Arthur 2014). They can be compared to Kepler's laws of planetary motion, discovered from the astronomical 'Big Data' collected by Tyge Ottesen Brahe (1546-1601, first critical edition by Rawlins 1993). Informal models produced by such macroscopic research methodologies (e.g., Schich et al. 2014) can be articulated using words, and are good as scaffolding for organizing thoughts. However, they have neither explanatory nor predictive powers, because we know what they are, but not why they are the way they are. To be able explain historical phenomena and predict when they will recur, we need the historical equivalent of Newton's laws. These are formal models, which can be stated either in equation form or as a set of rules (ABMs). To emulate how Newton's laws explain Kepler's laws, and understand the role of change in historical studies, Peter Turchin built top-down models describing how civilizations expand, through agriculture or military conquests (Turchin 2003). By extracting a few parameters from highly aggregated historical data, Turchin was able to show how closely his technology-driven *cliodynamics* follow the historical trajectories of the major civilizations in the world (Turchin and Nefedov 2009).

Encouraging as it may seem, *cliodynamics* overlook the role of human agency. Human societies always have the need to make decisions, whether it is to trade, to go to war, or to sue for peace. Unfortunately, if we follow the *cliodynamics* approach to its logical conclusion, no part of history would have turned out differently, i.e. history is inevitable. This is contrary to what Gordon Woo is theorizing in his calculation of catastrophes (2011). To put human agency back into history, and allow history to be contingent upon the decisions made (and therefore admit counterfactual histories), our ultimate goal remains the creation of historical ABMs.

3 Agent-Based Modelling and Simulations (ABMS)

Compared to equation-based modelling, which goes back as far as Newton in the seventeenth century, agent-based modelling and simulation (ABMS) has a very short history. While there may be early thinkers who contemplate the collective consequence of decisions made by many individuals, as a mode of inquiry ABMs can only be regarded to have started in the middle of the twentieth century. This is because the history of ABMS cannot be divorced from the development of the electronic computer. As early as 1971, Nobel Laureate in Economics Thomas Crombie Schelling developed a toy model of segregation, in which happy agents stay put while unhappy agents move (Shelling 1971). Agents are happy if more than a certain fraction of their neighbours are similar to themselves, and are unhappy otherwise. Using coins and graph paper to run the simulations, Schelling was able to show that any level of preference for neighbours similar to themselves will lead to segregated neighbourhoods.

In the 1980s, when the electronic computer was starting to become popular as a research tool in universities, political scientist Robert Axelrod hosted a tournament for computer programs to play the Prisoner's Dilemma against each other (Axelrod 1980). In this first true agent-based simulation, the agents were the computer programs that had to decide what strategy to use when playing against other computer programs who are also capable to deciding on or changing their strategies. Later in the 1980s, we also saw the development of ABMs called *Boids* by Craig Reynolds to explain the flocking of birds and the schooling of fishes (1986). In these models, the agents follow three simple rules: (1) move in the average direction of neighbours, (2) stay close to neighbours, and (3) avoid collisions with neighbours, and adjust their velocities accordingly.

Around the same time, computer scientist John Holland and economist Brian Arthur were also developing the world's first artificial stock market, where adaptive agents in the form of computer programs buy and sell stock according to their predictions of how the stock price will change (Palmer et al. 1994). In this Santa Fe Institute's Artificial Market model, agents trade with their best prediction model, out of a list of prediction models they maintain. These prediction models then evolve over time by random mutations or by mating between models. They found that the market and the agents never settle down, and are constantly generating booms and busts like in the real market. In 1991, John Holland and John Miller published a paper referring to their model as an 'agent-based model', and the name stuck (Holland and Miller 1991).

Since then, the field of ABMs expanded rapidly. While economists were the first adopters of this new computational methodology, ABMs quickly spread to other social sciences. In particular, Joshua Epstein and Robert Axtell created the *Sugarscape* ABM to explain the rise and fall of a large North American Indian settlement, and popularized ABM for social scientists by writing their book on this project (Epstein and Axtell 1996). As of now, ABM has become a fairly mature technology. There are now major conferences on ABM, and also summer/winter

schools on ABM attended by postdoctoral researchers and PhD students, taught by leading experts in the field. However, the spread of ABM as a tool has not been uniform across the social sciences and humanities, history being a late adopter. In this section, we will first describe how historians can build ABMs, what they can learn from ABMs, and how ABMs can help them transform history as a discipline.

3.1 Requirements and Prescriptions to Build Data-Driven Simulations

According to Cain (2014, 1), "a mathematical model is an attempt to describe a natural phenomenon quantitatively. Mathematical models in the molecular biosciences appear in a variety of ways: some models are deterministic while others are stochastic, some models regard time as a discrete quantity while others treat it as a continuous variable, and some models offer algebraic relationships between variables while others describe how those variables evolve over time". ABMs, though rule-based instead of equation-based, share many of the above characteristics. Historians wishing to reap these benefits must first learn how to build ABMs. To support the development of such models, there are specific requirements besides time and space that historians must take into consideration, when they construct their databases or re-structure them accordingly.

Firstly, in the ontology of the entities that are used in the database engineering, historians should identify:

- necessarily, agents, that are the entities, considered as individuals and/or collective wholes (i.e., governments, families, etc.), capable of setting goals, interact with other agents, and react to the environment and its changes;
- necessarily, events, from which they can extract the actions needed to achieve goals;
- possibly, conditions (due to the environment and/or other agents), that are the most important external factors influencing the decision-making process of the agents;
- possibly, preferences (built-in conditions), that are the arbitrary choices made by the agents to pursue their goals.

For example, in the Engineering Historical Memory (EHM) project, as first basic entities, besides time and space (always included), we decided to identify:

- governments and families as agents;
- trade, conflict, and diplomacy as filtering categories for actions like single treaties, embassies, travels from one place to another, buy/sell/loan/stockpiling of goods, battles, shipwrecks, sieges, wars, etc.;
- non-agent entities (goods, coins, ships, etc.) as conditions or preferences.

Secondly, the database needs to facilitate or at least allow for the retrieval of agent-action-condition (who did what and why) triplets, so that historians can

visualise the frequencies of actions taken under specific conditions by agents and how they depend on time and space.

We then codify the most frequent or most important agent-action-condition triplets as our rules for the ABM. If preferences can be inferred, these will also be included. Otherwise, we may—through consulting human experts—endow our agents with heterogeneous preferences consistent with behaviours of peoples of that time and space, as input parameters for our ABM. At this point, we are ready to write the computer program to simulate the ABM.

If what is missing in computational history is the macroscopic modelling, which grasp big data "through a process of compression, by selectively reducing complexity until once-obscure patterns and relationships become clear" (Graham et al. 2016, 1), we propose to fill this crucial gap following the *Annales* experience, and the consequent development of microhistory from the *histoire événementielle* (Le Goff and Nora 1974 and 1985, Burguière 2006 and 2009). In our vision, macroscopic models can be inferred by microhistory. In this perspective, big history is what emerges from *all* microhistories interactions.

Microhistory (Ginzburg 2012, 193-214) studies well-defined single historical units/events to ask—as defined by Charles Joyner—"large questions in small places" in contrast with large-scale structural views (Joyner 1999, 1). The most famous example being Carlo Ginzburg's *Il formaggio e i vermi* (in Italian 1976 and in English 1980). In the book, which is considered to have initiated this research field in historical studies, the author wrote: "The historians have long since learned that history is the history of men, not of the "great," and the closer you get up to everyday reality the better you decipher the past, and then grasp the sense of immediacy with the problems, the connections with today's present, i.e., history".

3.2 Under What Conditions Can We Learn from Big Simulations?

In his 2014 position paper, Michael Gavin argued for the adoption of ABM in history, as a means of encapsulating the complexity of historical events in terms of a small number of rules. Following Joshua M. Epstein and Robert L. Axtell (1996), Gavin calls this feature of ABM 'generative simplicity'. However, Gavin does not explain how ABMs are to be built starting from data. We explained in Section 3.1 how ABMs can be built in a data-driven process (which we want to promote). More importantly, it is not clear whether the small set of 'generative' rules are static, or whether they are adaptive and can change in time in response to evolutionary pressures. As John Holland had demonstrated, while the meta rules are the same (mutation, mating, selective reproduction), the rules themselves never settles down, and we have "perpetual novelty" in the system (1989, Introduction). Can we say we have an understanding of historical processes in terms of this ever-changing set of rules?

Also, when we simulate the ABMs, we end up with a large number of simulated histories. What then do we mean, when we say that we understand the observed history with the aid of simulated histories? To unpack this, let us suppose we understand much of human preference and behavior, but we do not have complete data on the population. After building an ABM, we would need to make assumptions on the preferences of the agents. Naturally, different assumptions will give us different simulation outcomes. However, if we believe that history can be 'understood', then the number of outcomes that are qualitatively different must be small compared to the number of assumptions we simulate. This means that a large number of simulations with different assumptions will give rise to the same qualitative outcome. There are a few inevitable results we can discuss.

First, some outcomes can emerge from a huge number of assumptions, whereas other outcomes appear only for a much smaller number of assumptions. We say that the former outcomes are robust, and the latter outcomes are fragile. Outcomes are not equal in this sense, and we can classify them using ABMS. Second, since many different assumptions give rise to the same qualitative outcome, many aspects of our assumptions must be unimportant (for otherwise they would have changed the simulation outcome). It may be that only a few aspects of the assumptions are important. This realization means that the outcome may be explained by a few key factors. Third, the historical trajectories leading to two qualitatively different (for example war versus peace) outcomes may follow each other closely until some point in time where they diverge. This point in time is when the simulated histories cross a tipping point. By comparing the key factors leading to the two different outcomes, and understand how they are different, we begin to have a better understanding of tipping points and regime shifts in history.

3.3 Tipping Points. A scalable solution to investigate change (i.e. the fundamental and nonlinear force of history)

Finally, to derive causal narratives of world history, and identify causative mechanisms and processes, we need to better understand what tipping points are (Gladwell 2000) and are not. In the discussion, above, we have already mentioned that a tipping point separates two qualitatively different set of historical trajectories. Certainly, a tipping point can be an event, and so the action associated with the event can be understood as a cause. However, let us make clear here that the action in the tipping point event is merely the cause of the event, but not the separation of historical trajectories. To understand this, we should think of 'the straw that broke the camel's back'. The laying of this straw onto the camel's back is clearly the tipping point, but it is no more causal than all the other strands of straw on the camel's back when it broke.

Ultimately, the causal narrative we would like to take away from this is that we have been adding load onto the camel, and thus drive the camel closer and closer to the tipping point. From the historical narrative extracted from the database, and the bundle of counterfactual trajectories that it is grouped with, the causal factor must be identified with the chain of key factors along these historical trajectories.

How then do we understand tipping points? Shortly after historical trajectories diverge, we can extract the chains of causal factors along each bundle of trajectories. We can compare these chains to identify the main difference in the chains of causal factors that lead to one bundle of trajectories going to one outcome, and another bundle of trajectories going to another outcome. This difference in causal factors, compared to the actions in the tipping point event, then tells us how small decisions that seemed inconsequential eventually turn out to have large impacts on the outcomes.

Finally, because we produced these counterfactual histories using a microscopic ABM, we can test in simulations what kind of changes to the preferences and behaviors of agents as the simulations is in progress will change the outcomes. Naively, if we have the preferences and behaviors of agents change continuously from the key factors of one outcome to the key factors of the other outcome, we should be able to change the outcomes for some of the simulations leading to an undesirable outcome.

4 Conclusions. Learning from Computational History

We repeatedly call for people to “learn from history”. *Historia vero testis temporum, lux veritatis, vita memoriae, magistra vitae, nuntia vetustatis, qua voce alia nisi oratoris immortalitati commendatur?* By what other voice, too, than that of the orator, is history, the evidence of time, the light of truth, the life of memory, the directress of life, the herald of antiquity, committed to immortality? (Marcus Tullius Cicero, *De Oratore*, II, 36). If we read this famous Cicero's quote through the lens of the thermodynamic paradigm, which holds that a perfect description of a given moment or set of conditions in history would provide a knowledge of future conditions—and assume that "the new society comes into being in the womb of the old” (Lechte 2003, 106), our increasingly complex world should cherish as much as possible the treasure of human experiences (*the* data), to increase resilience and sustainability and to nurture innovation (Nanetti et al. 2013, 104-105).

However, the circumstances surrounding historical episodes are never identical, and certainly not the same as the circumstances we find ourselves in. If war have been averted in the past because of certain diplomatic gambles, we may not be able to reuse them in the present day, because circumstances have changed, and also the actors have changed. Nevertheless, if we believe that history of our society is the result of selection between a small number of outcomes when it is presented with complex inputs, then ABM can help us understand the

relationships between different outcomes, and transitions between outcomes is only possible for neighbouring outcomes in some sort of phase diagram of outcomes. More importantly, ABM simulations can help us identify the key factors driving the simulations to a particular outcome, and what is the nature of the tipping point separating this outcome from a neighbouring one.

Computational analysis can borrow models from ecology, evolution, dynamical systems, and complexity theory (e.g., Holland 1989, 2000, 2012), and relate them to historical processes to extend the humanities capabilities and give new strength to the framework and prescriptions of century-old philological, historical, art historical, and anthropological methodologies. In doing so, we can aggregate knowledge for a better understanding of the present and transmit knowledge to influence the future values we desire more.

By producing counterfactual histories from ABM simulations and comparing these against the recorded histories, we can detect tipping points. If we are to learn from history and not make the same mistakes that our forebears did, we must understand the reasoning and decision processes that led to these tipping points. Only then can we acquire the wisdom to steer human civilisation towards desirable outcomes, and master the art of living together, with the consciousness of how false beliefs can change history (Eco 1998).

In our vision, this narrative-driven analysis of historical big data can lead to the development of multiple scale agent-based models, which can be simulated on a computer to generate ensembles of counterfactual histories that would deepen our understanding of how our actual history its related historiographies developed the way they did.

It entails the creation and advancement of databases (relational, graph, and hybrid), algorithms, computational, statistical, and complexity techniques and theories to solve formal and practical problems arising from the study, and the interpretation, conservation, and management of historical data and information.

In this way, the historians' major strength, their training in special units, can overcome its major weakness, the practical and ideological narrowness of specialised expertise, by adding to their data sets other historian's datasets and test their theories with multinational additional types of approaches that were not part of their training.

Acknowledgments This study has been funded by 2016 Microsoft Research Asia Collaborative Research Program and 2016 Microsoft Azure for Research. The research project, called Engineering Historical Memory (EHM), was first theorized by Andrea Nanetti in 2007, when he was Visiting Scholar at Princeton University. The actual web development initiated in 2012 when he was Visiting Professor at the University of Venice Ca' Foscari, and since 2013 has been carried out at Nanyang Technological University (NTU Singapore), where he is Associate Chair (Research) in the School of Art, Design and Media, and has been funded among others by an NTU Start-up Grant (2014-2016 M4081357), 2014 Microsoft Research Asia Collaborative Research Program, 2015 Microsoft Azure for Research, 2016 Microsoft Research Internship Program. This study contributed to the application and kick off of the NTU TIER 1 Grant (2017-2019) on "Data Consolidation for Interactive Global Histories (1205-1533) within the NTU National and International Research Network: Towards an NTU Interdisciplinary Laboratory for Data-Driven Agent-Based Modelling and Simulations for Historical Sciences". The domain of

EHM (www.engineeringhistoricalmemory.com) is administrated by Meduproject S.r.l., an Italian Pte Ltd. company established in 2002 by Andrea Nanetti as academic spin-off of the University of Bologna (Department of Histories and Methods for Cultural Heritage Conservation), after having been awarded in 2001 a prize in the first Italian business plan competition devoted to projects with high content of knowledge and having been financially supported by the Italian National Agency for New Technologies, Energy and Environment.

References

Abulafia, David. 2011. *The Great Sea: A Human History of the Mediterranean*. New York: Oxford University Press.

Ackoff, Russell L. 1989. From Data to Wisdom. *Journal of Applied Systems Analysis* 16: 3-9.

Aiden, Erez, and Jean-Baptiste Michel. 2013. *Uncharted: Big Data as a Lens on Human Culture*. New York: Riverhead.

Alvarez, Alejandro J., Carlos E. Sanz-Rodríguez, and Juan Luis Cabrera. 2015. Weighting Dissimilarities to Detect Communities in Networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 373.2056. doi:10.1098/rsta.2015.0108.

Axelrod, Robert. 1980. Effective Choice in the Prisoner's Dilemma. *The Journal of Conflict Resolution* 24/3:379-403.

Barthes. Roland. 1967. Le discours de l'histoire. *Social Science Information* 6/4:63-75.

Barthes. Roland. 1972. *Critical Essays*, transl. R. Howard. Evanston Northwestern University Press.

Big History Project. 2012-ongoing Big History Institute, Macquarie University. https://www.bighistoryproject.com. Accessed 15 January 2017.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast Unfolding of Communities in Large Networks, 2008. *Journal of Statistical Mechanics: Theory and Experiment* 2008. doi:10.1088/1742-5468/2008/10/P10008.

Brahe, Tycho. 1993. Tycho's Star Catalog: The First Critical Edition, ed. Dennis Rawlins. *DIO: The International Journal of Scientific History* 3.

Brian Arthur, William. 2009. *The Nature of Technology. What it is and how it Evolves*. New York: Free Press.

Brian Arthur, William. 2014. *Complexity and the Economy*. Oxford: Oxford University Press.

Brughmans, Tom, and Jeroen Poblome. 2016. MERCURY: an Agent-Based Model of Tableware Trade in the Roman East. *Journal of Artificial Societies and Social Simulation*. doi:10.18564/jasss.2953

Brughmans, Tom, and Jeroen Poblome. 2016. Roman bazaar or market economy? Explaining tableware distributions through computational modelling. *Antiquity*. doi:10.15184/aqy.2016.35

Burguière, André. 2006. *L'École des Annales: Une histoire intellectuelle*. Paris: Odile Jacob

Burguière, André. 2009. *Annales School: An Intellectual History*. Ithaca NY: Cornell University Press.

Cain, J.W. 2014. Mathematical Models in the Sciences. *Cellular and Molecular Life Sciences*. doi:10.1007/978-1-4614-6436-5_561-1

Calvino, Italo. 1988. *Lezioni americane*. Milano: Garzanti. [Calvino died during the night between the 18th and the 19th of September 1985. The first edition of these lectures was published posthumous in May 1988 as *Lezioni americane. Sei proposte per il prossimo millennio*, with an introductory note by Esther Calvino].

Calvino, Italo. 2006[24]. *Lezioni americane*. Milano: Oscar Mondadori.

Cheong, S.A., A. Nanetti, and M. Filippov. 2016. Digital Maps and Automatic Narratives for the Interactive Global Histories. *The Asian Review of World Histories* 4/1:83-123. doi:10.12773/arwh.2016.4.1.083.

CHIA. 2011-ongoing. http://www.chia.pitt.edu/. Accessed 15 January 2017.

Cohen, Daniel J, and Roy Rosenzweig. 2005. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press.

Cornwell, Benjamin. 2015. Social Sequence Analysis. Methods and Applications. New York: Cambridge University Press.

Digital Atlas of Roman and Medieval Civilizations. 2007-ongoing. Harvard University. http://darmc.harvard.edu. Accessed 15 January 2017.

Eco, Umberto. 1998. *Serendipities: Language and Lunacy*, transl. W. Weaver. New York: Columbia University Press.

Eddington, Arthur Stanley. 1929. *The Nature of the Physical World. The Gifford Lectures 1927*. New York: Macmillan. Cambridge UK: Cambridge University Press. See also the edition Annotated and Introduced by H. G. Callaway. Newcastle upon Tyne: Cambridge Scholars Publishing 2014.

Engineering Historical Memory. 2012-ongoing. Meduproject S.r.l. (spin-off company of the University of Bologna) and Microsoft Azure. http://www.engineeringhistoricalmemory.com. Accessed 15 January 2017.

Epstein, Joshua M., and Robert L. Axtell. 1996. *Growing Artificial Societies. Social Science from the Bottom Up*. Washington DC: Brookings Institution. Cambridge: and MIT Press.

Galasso, Giuseppe. 1984. Fonti storiche [Historical sources]. In *Enciclopedia del Novecento*, VII, 198-212. Roma: Istituto dell'Enciclopedia Italiana.

Galasso, Giuseppe. 2000. *Nient'altro che storia [Nothing but history]*. Bologna: Società Editrice Il Mulino.

Gavin, Michael. 2014. Agent-Based Modeling and Historical Simulation. *Digital Humanities Quarterly*. doi:digitalhumanities.org/dhq/vol/8/4/000195/000195.html#p2

Gell-Mann, Murray. 1997. The Simple and the Complex. In *Complexity, Global Politics, and National Security*, eds. David S. Alberts and Thomas J. Czerwinski, 2-12. Washington, DC: National Defense University.

Gilbert, Felix. 1990. *History: Politics or Culture? Reflections on Ranke and Burckhardt*. Princeton NJ: Princeton University Press.

Ginzburg, Carlo. 1976 (Italian) and 1980 (English). *Il formaggio e i vermi. Il cosmo di un mugnaio del '500*. Einaudi: Torino. In English: *The Cheese and the worms: the cosmos of a 16th-century miller*. Transl. J. Tedeschi and A. Tedeschi. Baltimore: Johns Hopkins University Press.

Ginzburg, Carlo 1986 (Italian) and 1989 (English). *Miti, emblemi, spie*. Torino Einaudi. In English: *Clues, Myths, and the Historical Method*. Transl. J. Tedeschi and A. Tedeschi. Baltimore: Johns Hopkins University Press.

Ginzburg, Carlo. 2001. Conversare con Orion. *Quaderni storici* 23/3:905-913.

Ginzburg, Carlo. 2006 (Italian) and 2012 (English). *Il filo e le tracce. Vero falso finto*. Bologna: Feltrinelli. In English: Threads and Traces. True False Fictive. Transl. A. Tedeschi and J. Tedeschi. Berkeley Los Angeles London: University of California Press.

Ginzburg, Carlo. 2012. Microhistory, two or three things that I know about it. In Idem 2012. *Op. cit*., 193–214.

Girvan, Michelle, and Mark E.J. Newman. 2002. Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Science of the United States of America* 99.12:7821-7826.

Gladwell, Malcolm T. 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. New York: Little Brown.

Gould, Stephen Jay. 1989. *Wonderful Life: The Burgess Shale and the Nature of History*. London: Hutchinson Radius, 1989.

Grafton, Anthony. 1994. The Footnote from de Thou to Ranke. In *Proof and Persuasion in History*, eds. Anthony Grafton and Suzanne L. Marchand. *History & Theory* 33/4:53–76.

Grafton, Anthony. 1995. *Die tragischen Ursprünge der deutschen Fußnote*. Berlin: Wagenbach.

Grafton, Anthony. 1997. *The Footnote: A Curious History*. Cambridge: Harvard University Press.

Grafton, Anthony, and Suzanne L. Marchand (eds.). 1994. Proof and Persuasion in History. *History & Theory* 33/4.

Grafton, Antony, Anja Goeing, and Paul Michel, and Adam Blauhut (eds.). 2013. *Collectors' Knowledge: What Is Kept, What Is Discarded/Aufbewahren oder wegwerfen: Wie Sammler entscheiden*. Leiden: Brill.

Graham, Shawn, Ian Milligan, and Scott Weingart (Eds.). 2016. *Exploring Big Historical Data. The Historian's Macroscope*. London: Imperial College Press.

Grinin, Leonid E. Grinin, and Andrey V. Korotayev (eds.). 2010. *History & Mathematics. Trends and Cycles*. Volgograd: 'Uchitel' Publishing House.

Gruber, T.R. 1993. A Translation Approach to Portable Ontologies. *Knowledge Acquisition* 5/2:199-220. doi:tomgruber.org/writing/ontolingua-kaj-1993.htm.

Gruber, T.R. 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies* 43/4-5:907–928. doi:tomgruber.org/writing/onto-design.htm

Guarino, Nicola, Daniel Oberle, and Steffen Staab. 2009. What is an *Ontology*. In *Handbook on Ontologies*, eds. Steffen Staab, and Rudi Studer, 153-176. Berlin-Heidelberg: Springer Verlag.

Guldi, Jo and David Armitage. 2014. *The History Manifesto*. Cambridge: Cambridge University Press. New updated version 5th February 2015, doi:http://dx.doi.org/10.1017/9781139923880. Accessed 7 January 2017.

Jockers, Matthew L. 2013. *Macroanalysis: Digital Methods & Literary History*. Urbana-Champaign: University of Illinois Press.

Joseph, Brian D., and Richard D. Janda. 2003. On Language, Change, and Language Change – Or, Of History, Linguistics, and Historical Linguistics. In *The Handbook of Historical Linguistics*, eds. Brian D. Joseph and Richard D. Janda. Oxford: Blackwell Publishing

Joyner, Charles W. 1999. *Shared Traditions: Southern History and Folk Culture*. Urbana: University of Illinois.

Halkin, Hillel. 2001. The Strange Adventures of Jacob d'Ancona: Is a memoir of China purportedly written by a 13th-century Jewish merchant authentic? And if not, what then? *Commentary Magazine* 111/4. doi:https://www.commentarymagazine.com/articles/the-strange-adventures-of-jacob-dancona

Hirschi, Caspar. *The Origins of Nationalism. An Alternative History from Ancient Rome to Early Modern Germany*. Cambridge UK: Cambridge University Press.

Hitzbleck, Kerstin, and Klara Hübner (Eds.). 2014. *Die Grenzen des Netzwerks 1200-1600*. Ostfildern: Jan Thorbecke Verlag der Schwabenverlag AG.

Holland, John. 1975. *Adaptation in Natural and Artificial Systems. An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor: University of Michigan Press.

Holland, John. 1989. *Induction: Processes of Inference, Learning, and Discovery*. Cambridge: MIT Press.

Holland, John. 2000. *Emergence: From Chaos to Order*. Oxford: Oxford University Press.

Holland, John. 2012. *Signals and Boundaries: Building Blocks for Complex Adaptive Systems*. Cambridge: MIT Press.

Holland, John H., and John H. Miller. 1991. Artificial Adaptive Agents in Economic Theory. *American Economic Review* 81/2:365-371.

Ladurie, Emmanuel Le Roy. 1973 and 1978. Le Territoire de L'Historien. 2 vols. Paris: Gallimard.

La Mothe Le Vayer, François de. 1669. *Œuvres*. 15 vols. Paris: Libraire Louis Billaine. [The fifth (and final) edition has been published in Dresde by Michel Groll between 1756 and 1759: *Œuvres de François de La Mothe Le Vayer, conseiller d'État ordinaire*, etc. Nouvelle

édition revue et augmentée. Imprimée à Pfœrten, et se trouve à Drede chez Michel Groell, 7 tomes en 14 volumes in-8].

Lechte, John. 2003. *Key Contemporary Concepts. From Abjection to Zeno's Paradox*. New York: SAGE Publications.

Le Goff, Jacques, and Pierre Nora (eds.). 1974. *Faire de l'histoire*. 3 vols. Paris: Gallimard.

Le Goff, Jacques, and Pierre Nora (eds.). 1985. Constructing the Past: Essays in Historical Methodology. Cambridge: Cambridge University Press. [This book presents a selection of ten of the most significant contributions to the three-volume *Faire de l'histoire*, 1974].

Manning, Patrick. 2013. *Big Data in History*. Basingstoke UK: Palgrave Macmillan.

Manning, Patrick. 2015. A World-Historical Data Resource: The Need is Now. *Journal of World-Historical Information* 2-3/2:1-6.

Mitchell, Melanie. 1996. An Introduction to Genetic Algorithms. Cambridge: MIT Press.

Momigliano, Arnaldo. 1985. History between Medicine and Rhetoric. *Annali della Scuola Normale Superiore di Pisa. Classe di Lettere e Filosofia. Serie III* XV/3:767-780.

Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for Literary History*. London New York: Verso.

Nanetti, Andrea. 2010. *Il Codice Morosini: Il Mondo Visto da Venezia (1094-1433) [The Morosini Codex: The World as Seen from Venice (1094-1433)],* 4 vols. Spoleto: Foundation CISAM.

Nanetti, Andrea, Siew Ann Cheong, and Mikhail Filippov. 2013. Interactive Global Histories: For a New Information Environment to Increase the Understanding of Historical Processes. In *2013 Culture and Computing*, 104-110. Los Alamitos: IEEE Computer Society.

Nanetti, A., and S.A. Cheong. 2016. The World as Seen from Venice (1205-1533) as a Case Study of Scalable Web-based Automatic Narratives for Interactive Global Histories. *The Asian Review of World Histories* 4/1:3-34. doi:10.12773/arwh.2016.4.1.003.

Nanetti, A., C.-Y. Lin, and S.A. Cheong. 2016. Provenance and Validation from the Humanities to Automatic Acquisition of Semantic Knowledge and Machine Reading for News and Historical Sources Indexing/Summary. *The Asian Review of World Histories* 4/1:125-132. doi:10.12773/arwh.2016.4.1.125.

Olstein, Diego. 2015. *Thinking History Globally*. Houndmills UK New York: Palgrave Macmillan.

Online Catasto of Florence, 1427-1429. 1969. Brown University, eds. David Herlihy, Christiane Klapisch-Zuber, R. Burr Litchfield, and Anthony Molho. http://cds.library.brown.edu/projects/catasto/overview.html. Accessed 2 January 2017.

Orlandini, Giovanni. 1913. *Origine del Teatro Malibran. La Casa dei Polo e la Corte del Milion*. Nozze Alverà-Trevisanato. Venezia.

Orlandini, Giovanni. 1926. Marco Polo e la sua famiglia. *Archivio veneto-tridentino* 9-10:1-68.

Palmer, Richard G., William Brian Arthur, John H. Holland, and Blake LeBaron and Paul Tayler. 1994. Artificial economic life: A simple model of a stock market, *Physica D* 75:264-274. https://www.phy.duke.edu/~palmer/papers/arob98.pdf. Accessed 15 January 2017.

Paolucci, Mario, and Stefano Picascia. 2011. Enhancing collective filtering with causal representation. In *2011 Culture and Computing*, 135-136. Los Alamitos: IEEE Computer Society.

Pavlus, J. 2015. A New Map Traces the Limits of Computation. A major advance reveals deep connections between the classes of problems that computers can—and [yet?] can't—possibly do. *Quanta Magazine*. doi: quantamagazine.org/20150929-edit-distance-computational-complexity.

Pelagios. 2011-ongoing. Pelagios Commons. http://commons.pelagios.org. Accessed 15 January 2017.

Pessoa, Osvaldo Jr. 2001. Counterfactual Histories: The Beginning of Quantum Physics. *Philosophy of Science* 68:519-530.

Popper, Karl. 1994. *Alles Leben ist Problemlösen*. München: Piper Verlag.

Popper, Karl. 1999. *All Life is Problem Solving*. Trans. P. Camiller. London and New York: Routledge.

Progetto Cronache Veneziane e Ravennati. Fondazione Casa di Oriani. http://www.cronachevenezianeravennati.it. Accessed 15 January 2017.

Radick, Gregory. 2005. Other Histories, Other Biologies. In Anthony O'Hear (ed.), *Philosophy, Biology, and Life*. Cambridge: Cambridge University Press, 21-47.

Reynolds, Craig. 1986. *Boids* [an artificial live program which simulates flocking birds in 2D]. http://www.red3d.com/cwr/boids. Accessed 15 January 2017.

Richthofen, Ferdinand von. 1876. Über den Seeverkehr nach und von China im Altertum und Mittelalter. *Verhandlungen der Gesellschaft für Erdkunde zu Berlin* 1876:86-97.

Richthofen, Ferdinand von. 1877. Über die zentralasiatischen Seidenstrassen bis zum 2. Jh. n. Chr. *Verhandlungen der Gesellschaft für Erdkunde zu Berlin* 1877:96-122.

Richthofen, Ferdinand von. 1877-1912. *China. Ergebnisse eigener Reisen und darauf gegründeter Studien*. 5 vols. Berlin: Reimer.

Robertson, Douglas S. 1998. *The New Renaissance: Computers and the Next Level of Civilization*. Oxford: Oxford University Press.

Robertson, Douglas S. 2003. *Phase Change: The Computer Revolution in Science and Mathematics*. Oxford: Oxford University Press.

Rosnay, Joël de. 1979. *The Macroscope: A New World Scientific System*. New York: Harper & Row.

Sanudo il Giovane, Marin. 1879-1902. *Diarii*, 58 vols., ed. Rinaldo Fulin, Federico Stefani, Nicolò Barozzi, Guglielmo Berchet, Marco Allegri. Venezia: Stabilimento Visentini cav. Federico Editore.

Schäfer, Wolf. 2001. Global Civilization and Local Cultures. A Crude Look at the Whole. *International Sociology* 16/3:301-319. Also in *Rethinking Civilizational Analysis*, eds. Saïd Amir Arjomand and Edward A. Tiryakian, 71-86. London: SAGE Publications.

Shelling, Thomas C. 1971. Dynamic Models of Segregation. *Journal of Mathematical Sociology* 1/2:143-186.

Schich, M., C. Song, Y.-Y. Ahn, A. Mirksky, M. Martino, A.-L. Barabási, and D. Helbing. 2014. A network framework of cultural history. *Science* 345:558–562. doi:10.1126/science.1240064.

Seshat: Global History Databank. 2011-ongoing. The Evolution Institute. http://seshatdatabank.info. Accessed 15 January 2017.

Spiegel, Gabrielle M. 1997. *The Past as Text*. Baltimore and London: The Johns Hopkins University Press.

Toynbee, Arnold J. 1934-1961. *A Study of History*, 12 vols. London: Oxford University Press.

Trismegistos. 2004-ongoing. University of Leuven and University of Cologne. http://www.trismegistos.org. Accessed 15 January 2017.

Turchin, Peter. 2003. *Historical Dynamics: Why States Rise and Fall*. Princeton: Princeton University Press.

Turchin, Peter, and Sergey A. Nefedov. 2009. *Secular Cycles*. Princeton: Princeton University Press.

Vendrix, Philippe. 1997. Cognitive sciences and historical sciences in music: Ways towards conciliation. In *Perception and Cognition of Music*, eds. Irène Deliège and John Sloboda, 64-74. Hove UK: Psychology Press.

Vlastos, Gregory. 1983. The Socratic Elenchus. *Oxford Studies in Ancient Philosophy* 1:27–58.

Waugh, Daniel C. 2007. Richthofen's "Silk Roads": Toward the Archaeology of a Concept'. *The Silk Road* 5/1:1-10.

Williams, Bernard. 2006. *The Sense of the Past*. Princeton: Princeton University Press.

Wang, Gungwu. 2016. Heritage and History. *3rd Singapore Heritage Science Conference* (Nanyang Technoogical University, Singapore, 25-25 January 2016). doi:youtube.com/watch?v=-0wXSnqcAlM&list=PLasWJveXPWTE02EHZ2zsxzPxU6m3-GQli&index=3. Accessed 6 January 2017.

Wong, Sylvia C., Simon Miles, Weijian Fang, Paul Groth, and Luc Moreau. 2005. Provenance-Based Validation of E-Science Experiments. In *The Semantic Web – ISWC 2005. Proceedings of the 4th International Semantic Web Conference (Galway, Ireland, November 6-10, 2005)*, eds. Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen, 801-815. Berlin-Heidelberg: Springer Verlag.

Woo, Gordon. 2011. *Calculating Catastrophe*. London: Imperial College Press.

Yule, Henry, and Henri Cordier. 1913–1916. *Cathay and the Way Thither*. 4 vols. London: The Hakluyt Society.