Measuring Cultural Transmission at Archaeological Scales: How Can We Improve Empirical Sufficiency?

Mark Ernest Madsen

A dissertation submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

University of Washington
2020

Reading Committee:
James K. Feathers, co-Chair
Benjamin Marwick, co-Chair
Carl P. Lipo, Binghamton University

Program Authorized to Offer Degree:
Anthropology

University of Washington

Abstract

Measuring Cultural Transmission at Archaeological Scales: How Can We Improve Empirical Sufficiency?

Mark Ernest Madsen

Co-Chairs of the Supervisory Committee:
Research Associate Professor James K. Feathers
Associate Professor Benjamin Marwick
Anthropology

Cultural transmission has long been a key organizing principle within anthropology, but the effort to formalize cultural transmission models and fit them to archaeological data is more recent, stimulated by work by Robert Dunnell in the 1970's. Since then, the use of cultural transmission modeling in archaeology has branched into several research programs: one macroevolutionary, employing phylogenetic methods; and one microevolutionary, employing models derived from population genetics. A third research program, focused on intermediate or ”mesoscopic” scales and seriation as a finer-grained counterpart to phylogenetic and cladistics, is being developed by Carl Lipo and the present author.

This dissertation collects research papers by the author since 2012 which examine two questions. First, are equifinality issues encountered in the microevolutionary research program solvable or do they prevent us from employing individual-scale models? Second, to the extent that equifinality cannot be circumvented, can we construct better approaches at the mesosopic scale appropriate to coarse grained, time averaged data?

Two papers examine the first question, using simulation modeling and statistical methods to test whether theoretical models can be distinguished even in principle. The first paper examines the effects of temporal aggregation, which is ubiquitous in the archaeological record, on our ability to distinguish between cultural transmission models, and finds significant issues in doing so with time averaged data. The second paper examines the effects of population heterogeneity in social learning modes, which is well documented from living human and animal populations. I find that heterogeneous mixtures of social learning rules can be

identified statistically, but only with synchronic censusing of the population, and that time averaging and small samples render mixtures indistinguishable from pure unbiased copying.

Turning to the second question, three papers continue my long-term research into reshaping the classical seriation method into a tool for tracing the structure of cultural transmision at regional scales. One short paper examines the combinatorial structure of the seriation problem when we admit multiple subsolutions. A second paper seeks to increase the size of possible seriations, which is necessary to incorporate significant spatial variation and yield a tool usable for investigating the history of cultural transmission in a region. We increase the potential size of seriation solutions by switching from unimodality to distance minimization as the ordering criterion, yielding "continuity" seriation as a distinct method. A third paper in this group then applies continuity seriation graphs as the observable variable, in a methodological study of how to construct models of how cultural transmission was structured at the regional scale. This paper introduces "interval temporal networks" as a way to formalize our hypotheses about regional interaction and transmission, and explores a statistical method for summarizing the topology of seriation graphs, to assess their fit to our regional interaction models.

A final paper examines a different kind of mesoscale question: how do we begin to model not just the spatiotemporal structure of past cultural transmission, but its *content* as well. The chapter models the dependency structure of the knowledge required to construct complex artifact types, through the "prerequisites" needed for each step, and introduces a model where transmission of subsequent traits requires learning their prerequisites first. This simplified model of "structured" cultural traits is then used to explore the "learning hypothesis" for behavioral modernity, by looking at the richness and depth of knowledge gained when transmission is mostly accomplished by simple imitation compared to learning via a teacher. The results are suggestive that the learning hypothesis can account for the increased richness of "behaviorally modern" hominids, and more importantly, points the way to more substantive and technologically informed cultural transmission models.

Contents

List of Figures

List of Tables

Acknowledgements

The University of Washington Department of Anthropology, with whom I have been fortunate to be associated in one form or another since 1985, provided financial support, lab space, computer time (back in the days when you paid by the hour!), and other resources needed to conduct this research. Most of all, the Department tolerated and supported the unusual and circuitous path my graduate studies and research have taken. Thank you for your support. I was introduced to the Department by Stevan Harrell through his participation in the UW Honors Program. Faculty with whom I had significant interaction and who influenced me include Jim Green, David Spain, Stevan Harrell, Miriam Kahn, Laura Newell, Steven Goodreau, Gerald Eck, Eric Alden Smith, Julie K. Stein, Donald K. Grayson and Robert Wenke. Thank you all for a superb education. I also wish to thank Peter D. Ward, whose graduate paleobiology course in the spring quarter of 1987 substantially increased my understanding of evolutionary biology and broadened my research horizons.

I thank everyone who has served on both incarnations of my dissertation committee, including Robert C. Dunnell, Donald K. Grayson, Thomas Stoebe, Carl Bergstrom, Carl Lipo, Ben Marwick, and James K. Feathers. In particular, Jim Feathers and Ben Marwick provided strong encouragement at a key moment, and my gratitude for that push is eternal. Thank you both.

Carl Lipo has been a friend, colleague, and much more since 1988. Our collaboration is strongly represented in these pages, and I look forward to many more years of thinking about these issues and collaborative work together.

Over the years, my colleagues from the Department have become friends, many lifelong and beyond. I want to specifically thank Sarah Sterling, Fran Hamilton, Michael Pfeffer, Betsy Scharf, Steve Cole, and Chris Pierce. Most especially, Kim Kornbacher left us before I could finish this and share the joy with her, but she lives forever in my heart. Kris Wilhelmsen—please know that you and Kim are forever family to me!

A very special thanks goes to Lara Braithwaite, life partner and friend, for sustaining me through the latter stages of this project with great love, encouragement, and steadfast support. Thank you for our life together!

Finally, I thank Robert Chester Dunnell for the gift of his teaching, his thought, his insatiable curiosity, and his discipline. Dr. Dunnell was the very model of a scientist and scholar, and his influence can be seen on every page. From the moment I walked into his Archy 497 classroom in the autumn of 1986, he has been an inspiration, and set a standard to which I continue to try to attain.

My mother, Joy Colleen Berkey Madsen (1944-2005), encouraged me from the earliest age in a love of learning, not merely because it would lead to opportunity, but because knowledge and learning were important values in and of themselves. Her own opportunities for higher education were limited by the need to support herself from a young age, and then by the need to work and support her family, but she supported my education in every way possible.

This dissertation is both dedicated to her, and many ways a joint accomplishment.

CHAPTER 1

Introduction and Research Problem

1.1 Introduction

The study of human behavior within a scientific, Darwinian framework is no longer an upstart enterprise, operating on the fringes of biology and several social sciences such as anthropology, psychology, and economics. Only anthropology, however, can rightly claim responsibility for introducing and elaborating the idea of the "transmission" of culture, and it is the only social science for which cultural transmission is—and has been, for more than a century—a central organizing concept for the discipline (Lyman, 2008). Over the past 40 years, intense interdisciplinary research in anthropology and related fields has yielded a large cohort of researchers pursuing a range of Darwinian based investigations into human behavior. This cohort includes those conducting studies using gene-culture coevolution (or "cultural transmission theory" more broadly) (e.g., Boyd and Richerson, 1985; Cavalli-Sforza and Feldman, 1981; Mesoudi et al., 2006b; Richerson and Boyd, 2005), social psychological theories of norms and social epistemology (e.g., Binmore, 2005; Fehr and Fischbacher, 2004; Lewis, 1969), and the study of strategic social interaction using the tools of evolutionary game theory (e.g., Gintis et al., 2000, 2005; Gintis, 2014; Weibull, 1997). Taken together, these approaches offer increasingly productive means for understanding the structure and patterns of human behavior in areas such as cumulative cultural evolution and the evolution of cooperation.

These approaches to human behavior, however, tend to focus on the mechanisms by which information is passed and transformed between groups of people or from individual to individual. In the context of historical science, these provide proximate explanations. Tracing ultimate causation in order to account for why patterns of traits appear in varying frequencies through time and across space requires a way to document "descent with modification." Evolutionary archaeology, in particular, seeks to document evolutionary change from the archaeological record by extracting data on the transmission and inheritance of cultural traits.

For much of its history, anthropologists have traced and explained cultural patterns using intuitive, common-sensical methods (e.g., Lyman, 2009; Lyman and O'Brien, 2000a,b; O'Brien and Lyman, 1999b), which came to be partially systematized in the 1930's as "culture history." (Lyman et al., 1997; Lyman and O'Brien, 2001; Lyman and Michael, 2003; Lyman, 2008; O'Brien and Lyman, 1998; O'Brien et al., 2000). In the 1960's, a move to build explicit evolutionary explanations failed to capture the potential of the Darwinian paradigm, instead adopting a vitalistic and Lamarckian account of change (Dunnell, 1980). The combination of intuitive methods combined with theoretical models that offered only generalizations about change limited the success of anthropological efforts to explain human behavior using an evolutionary framework for decades.

Dunnell's influential works in the late 1970's and 1980s introduced a radical alternative to the "cultural evolution" then prevalent within archaeology (Dunnell, 1978, 1982, 1980, 1989). These works ultimately have supplied the basis and rationale for at least three distinctive research programs that focus on cultural transmission, each concentrating on a different scale of analysis. While there are other possible divisions between research programs, such as the focus on learning theory that comprises much of the work conducted in Japan (e.g., Feldman et al., 1996; Aoki and Feldman, 1987;

Aoki et al., 2011; Aoki, 2013, 2015; Nakahashi, 2013; Nishiaki et al., 2013; Terashima, 2013; Wakano and Aoki, 2007; Wakano et al., 2004), the division used here is focused upon issues of spatiotemporal scale and analytical level *sensu* Dunnell (1971).

The first program centers on an argument made by Dunnell (1978) that links the intuitive foundation of "culture history" with the large-scale needs of an evolutionary archaeology. This research area focuses on the separation of "homologous" similarity from similarity due to convergent adaptation in order to understand ancestor-descendant relationships. Its most notable area of development has been the adoption and development of cladistic methods borrowed from biology (Borgerhoff Mulder et al., 2006; Lyman et al., 1997; Lyman and O'Brien, 2006a; O'Brien and Lyman, 1999a; O'Brien et al., 2000, 2001, 2003; O'Brien and Lyman, 2003; O'Brian and Lyman, 2000; Prentiss and Laue, 2019; Prentiss et al., 2015; Tëmkin and Eldredge, 2007). This research program is "macroevolutionary" in flavor and combines methods from biogeography, demography, phylogenetics, and the comparative method to understand large-scale evolutionary history.

A second research program aims to use formal models of cultural transmission to explain the distributional characteristics of stylistic variation in artifact assemblages. This approach is "microevolutionary" in scale since it focuses on model fitting and inference within single assemblages or small sets of assemblages, taken to represent a population. The work started from Dunnell's (1978) in which differences were not subject to natural selection (i.e., "neutral traits") but also includes the pioneering modeling work of Boyd and Richerson (1985) and Cavalli-Sforza and Feldman (1981). Fraser Neiman (1995) provided a quantitative basis for this effort using the Wright-Fisher model from theoretical population genetics to derive predictions about artifact class diversity measures. Using these predictions one can evaluate whether these measures meet the expectations of neutrality. In this way, Neiman provided a practical test for determining whether sets of classes used to describe assemblages displayed signs of neutrality and thus are usable for tracing homology and evolutionary theory (Eerkens et al., 2006; Eerkens and Lipo, 2007; Lipo et al., 1997; Lipo and Madsen, 2001; Lipo, 2006).

The goals of this research program are several (Marwick, 2005). First, in addition to statistical testing for goodness of fit to the expectations of neutrality, researchers examined the conditions that lead to neutrality and the potential for selective pressures to be involved in the evolution of cultural traits (Bettinger and Eerkens, 1999; Bettinger, 2008; Eerkens and Lipo, 2005; Evans and Giometto, 2011; Pfeffer, 2001; Steele et al., 2010; Wilhelmsen, 2001). Second, a large group of researchers have been interested in employing Boyd and Richerson's (1985) models of transmission biases to characterize ways in which past populations may have had propensities for novelty-seeking, a bias towards conformity, or prestige-biased imitation (Acerbi and Bentley, 2014; Bentley and Maschner, 2001; Hahn and Bentley, 2003; Bentley and Shennan, 2003; Bentley et al., 2004, 2007; Bettinger and Eerkens, 1999; Herzog et al., 2004; Kohler et al., 2004; Mesoudi and Lycett, 2009; Shennan and Wilkinson, 2001; Shennan and Bentley, 2008). To the extent that this research tradition has made use of neutral models as a "null hypothesis" against which to test for departures, such studies have employed the same models and methods. For example, both types of work have tended to focus on examination of frequency patterns within an assemblage or small set of assemblages (see the detailed reviews by Kandler and Crema, 2019; Walsh et al., 2019).

A third research program is focused on scales in between single populations and the large-scale viewpoint of macroevolutionary studies. Carl Lipo and myself, in collaboration with Dunnell prior to his passing in 2010, have been engaged in exploring "mesoscale" methods for tracing transmission patterns within and across regions, using observable units which incorporate time and change directly. A focus above the "microevolutionary," we believe, is essential given the diachronic and time averaged nature of the archaeological record. Furthermore, the observable units we employ are critical to rendering our cultural transmission models empirically sufficient. Much of our work has involved extending classical seriation methods to be general purpose tools for constructing models of evolutionary histories (Lipo et al., 1997; Lipo, 2001b; Lipo and Madsen, 2001; Lipo, 2005; Lipo et al., 2015; Lipo and Madsen, 1997; Lipo et al., 1995; Madsen et al., 2008; Madsen and Lipo, 2014, 2015). Others who have focused on the mesoscopic include Kandler and Shennan (2013)'s important work on non-equilibrium neutral models, and Kandler's subsequent work with approximate Bayesian generative approaches (Kandler and Shennan, 2015; Kandler and Powell, 2018; Wilder and Kandler, 2015).

All three programs have had varying levels of success in the overall goal of using Darwinian evolutionary theory to account for cultural aspects of human behavior. The macroevolutionary research has demonstrated its utility for examining homology and tracing evolutionary relationships. The microevolutionary research program, on the other hand, has been struggling for the last decade to demonstrate its ability to fit detailed models of transmission bias to archaeological data (Kandler and Crema, 2019; Richerson and Boyd, 2008). Thus, much attention has been given to identifying sources of equifinality and seeking to remedy or "correct" for them (Barrett, 2019; Premo, 2010). In contrast, the mesoscopic work has sought to avoid problems of equifinality by focusing on an analytical level (i.e., multiple assemblages) that are a better match for the scale of a given empirical data set (e.g., Eerkens and Lipo, 2007; Lipo, 2001a; O'Brien et al., 2015; Perreault, 2019).

The issue of equifinality remains a serious challenge to the goal of building a fully evolutionary archaeology. To move forward, we need to ask some difficult questions. While we can match data to the expectations we derive from models, how do we know that other matches are not likely? Is it possible to "correct" for equifinality? Do studies at scales above single assemblages offer the best solution to this problem?

Based on the fundamental issue of equifinality, this dissertation has two parts. First, I explore whether the challenges faced by the microevolutionary research program are solvable ones and whether there are fundamental limitations on our ability to distinguish the details of cultural transmission from the archaeological record. Second, and following the mesoscopic approach, I investigate whether there are ways of constructing better observable units from our observations of the record that allow us to map homology and evolutionary history in empirically sufficient ways. These two questions have motivated my research for more than a decade, going back to joint work with Robert Dunnell, Carl Lipo, and Tim Hunt (Lipo et al., 1997; Lipo and Madsen, 1997; Lipo et al., 1995).

In the next section, I review the issues that have been discovered within the microevolutionary program in more detail. I describe conclusions reached from several pieces of my own research, included in this dissertation that aimed at understanding the limitations of the dominant approach to detailed, microevolutionary modeling in archaeology. Those limitations largely stem from the conceptual approach that early attempts at microevolutionary modeling took in archaeology. In particular, they are derived from the implicit willingness to ignore the mismatch between synchronic methods and modeling approaches, and the diachronic nature of the archaeological record. As I show in one of the papers included here, we are unlikely to simply "correct" for the diachronic nature of the records simply by creating better methods. Microevolutionary approaches simply require data that are not typically available in the archaeological record except in a few idiosyncratic cases (Scholnick, 2010; Mallios, 2014).

The idea that the initial promise of the microevolutionary program may be limited by fundamental equifinality issues will be disappointing to some. In particular, we may not be able to uniquely fit social psychological models of transmission to archaeological data, and this limits our ability to consider cultural transmission modeling as a bridge between social theory and most archaeological data. My own contention is that we should not expect such a bridge to be possible, given the nature of the empirical record we study (see also Perreault 2019). As Richerson and Boyd (2008) note, the archaeological record "speaks softly" on too many of the pieces of information one needs in order to make the microevolutionary modeling approach more than an interpretive heuristic in most cases.

At the other end of the spectrum, the macroevolutionary approach and cladistic modeling are important tools for mapping large-scale evolutionary history, but the method typically relies on presence and absence data for classes and types. This reliance means that macroevolutionary methods are limited in their ability to resolve detail in evolutionary history. To map homology and history at more detailed temporal and spatial scales, we need tools that can make use of our quantitative understanding of variation over time and space.

Instead, I argue that we need to build a distinct approach focused at "mesoscale" phenomena. By developing models and methods for scales between the individual level and large-scale phylogeny, we will hope to address the detailed record of cultural evolution in a region. The goal of building distinctive methods for the mesoscale is to gain substantial increases in empirical sufficiency, and build tools that inherently incorporate the diachronic nature of our data.

Based on this reasoning, the heart of my dissertation research addresses the means for constructing observable units from our observations of the record that allow us to map homology and evolutionary history in empirically sufficient ways. I explore how we can use data structures—–seriations and dependency graphs—–as the observational units or "features" (in data science terminology) for fitting models to archaeological data. This work combines aspects of "feature engineering" from machine learning and data science, where variables and data are combined in ways that provide the maximum ability to discriminate between hypotheses or models, and good old-fashioned unit construction using tools like seriation that have long history within our discipline.

1.2 Attempts to Assess Equifinality in the Microevolutionary Program

While early and paradigm setting efforts began in the late 1970's (e.g., Dunnell, 1978), the microevolutionary program was first established in the early work of Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985). These contributions were followed by a series of works in the late 1980s and 1990's (e.g., Dunnell, 1989; Neiman, 1990, 1995). The greatest growth of microevolutionary approaches, however, occurred after 2000 (e.g., Eerkens and Lipo, 2005; Hamilton and Buchanan, 2009; Kandler and Shennan, 2013, 2015; Jordan and Shennan, 2003; Shennan and Wilkinson, 2001; Perreault and Brantingham, 2011; Scholnick, 2010; Rorabaugh, 2014; Wilder and Kandler, 2015). Since that time, much attention has been given to the application of various statistical tests to a variety of data sets, both contemporary (e.g., Herzog et al., 2004; Hahn and Bentley, 2003), and archaeological (e.g., Jordan and Shennan, 2003; Mesoudi and O'Brien, 2008a; Shennan and Wilkinson, 2001). The goal of these studies has been to demonstrate the potential to examine modes of transmission that appear at the scale of individuals. Much of this work shares a common conceptual structure:

- A chosen model of transmission bias (or models) is compared to a model lacking bias (typically, the Wright-Fisher model of genetic drift);
- Predictions for a diversity statistic or the shape of a frequency distribution at equilibrium are made from all of the models, sometimes using analytic equations (from Wright-Fisher), but more often by simulation;
- An empirical data set of artifact class frequencies are compared to model predictions to see which model has the closest match.

Most of the early studies presented their results as seemingly clear-cut and implied that it was possible to differentiate between models given existing, previously-collected data on artifact classes. Within the last decade, however, the early clarity in the results has receded, especially after researchers began to reanalyze data sets using different approaches, with divergent results.

The European Neolithic Merzbach *Linearbandkeramik* (LBK) ceramic dataset has often been held as an example of how it is possible to isolate microevolutionary mechanisms. Kandler and colleagues (2015), however, note that after four studies previous to their own, the results remain conflicted as some studies support the hypothesis of neutrality for ceramic styles in the Merzbach LBK, while others reject neutrality in favor of anti-confirmist or novelty-seeking models of transmission. These analyses included a variety of methods that include variants of the diversity index method (Shennan and Wilkinson, 2001; Shennan and Bentley, 2008), power law fits (Bentley and Shennan, 2003; Shennan and Bentley, 2008), Kandler's non-equilibrium assemblage comparison method (Kandler and Shennan, 2013), and finally, an approximate Bayesian "generative" modeling approach (Kandler and Shennan, 2015). Each analysis offered compelling evidence about microevolutionary mechanisms, though the results varied on a study-by-study basis.

The failure to consistently replicate the initial conclusions drawn from the Merzbach LBK assemblages is not due to a faulty analytical method. Rather, it is due to the inherent problem in seeking to isolate individual-scale mechanisms from these data. In the past decade, this conclusion has become clearer when researchers began to focus on the sources of equifinality that might cause one to be unable to distinguish between biased and unbiased transmission models (Premo, 2010). Equifinality is a consideration whenever one seeks to account for samples of data from a complex empirical phenomenon, complete with chaos and nonlinearities (Bertalanffy, 1969). When our theoretical models are relatively simple and stochastic in nature, it is likely that multiple models can generate the same outcomes.

Equifinality has long been understood as an issue in archaeological interpretation and analysis (e.g., Gifford-Gonzalez, 1991; Kandler and Powell, 2018; Lamberg-Karlovsky, 1970; Lyman, 2004; O'Brien et al., 1998; Premo, 2010; Rafferty et al., 2008). It has not, however, been the subject of systematic study unlike disciplines such as geomorphology, climatology, and especially hydrology (e.g., Culling, 1987; Beven, 1996; Cicchetti and Rogosch, 1996; Aronica et al., 1998; Savenije, 2001; Beven, 2006; Ebel and Loague, 2006; Bonham et al., 2009; Vrugt et al., 2009; Cruslock et al., 2010; Khatami et al., 2017, 2019). The lack of attention in archaeology on issues of equifinality has long thwarted progress towards consistent method development and cumulative knowledge generation. And the problem is one that can be addressed: given that varying factors can cause different transmission models to yield similar outcomes, it is incumbent upon us to design better analytical models and methods that are designed to circumvent the problem.

The most readily apparent contributor to equifinality in our models is the mismatch between the synchronic structure of our models and predictions, and the diachronic, aggregate nature of the archaeological record. While many evolutionary models address the structure of variability generation at particular points in time, the data we evaluate represent a cumulative set of events of varying duration. Thus, a key step in any evolutionary model is the derivation of model predictions for distributional characteristics or summary statistics that can be used to compare with our class frequency data. In most cases, however, the predictions or test statistics are synchronic; that is, they describe the situation that obtains in a hypothetical population subject to the transmission model at a point in time. Much of classical population genetic theory is structured to describe conditions at points in time. For example, in pre-genomic population genetics researchers used stochastic models of genetic sampling within a population to produce predictions from the stationary distributions of the stochastic process and the extraction of marginal distributions or various statistics about the population or samples from the population (e.g., Ewens, 1972, 2004; Slatkin, 1994; Watterson, 1974, 1978).[1] Given the fact that the archaeological record represents the accumulation of events over time, naively borrowing this conceptual structure from population genetics without serious modification has been a major mistake. The lack of modifications to account for archaeological data is compounded by our inability to fully parameterize cultural transmission models, as Richerson and Boyd (2008, 301-302) noted in their critique of microevolutionary efforts in archaeology. If our models rely on parameters such as population size but we cannot directly measure population using the archaeological record, we already working from a vastly weakened position.

The archaeological record, however, is diachronic. As a result, the nature of the archaeological record has fundamental consequences for cultural transmission modeling. It is clear to all archaeologists that the archaeological record is not a sequence of "moments in time" but rather a cumulative record of artifact deposition whose temporal properties depend not just upon the intensity of use, but upon the sedimentary and geomorphological context (Schiffer, 1983, 1987; Stein, 2001, 1987, 1993, 2001; Stein et al., 2003). As a result, there is a growing understanding that many, if not most, of our samples of the archaeological record reflect deposition over variable and significant spans of time. This fact means that archaeological data—counts and frequencies of artifact types, species in faunal assemblages and skeletal part inventories, paleobotanical assemblages, indeed, every kind of archaeological data—are potentially "time-averaged" (Walker and Bambach, 1971). As a result, our data almost never refer to a specific configuration of a population, but are a kind of aggregate observation over a duration.

The effects of "time averaging" have been studied in a variety of contexts within archaeology. The most prominent studies tend to be in Paleolithic deposits and certain depositional contexts such as aeolian environments and surface contexts that are comprised of stable and old surface ages (e.g., Bai-

---

[1] This style of modeling and analysis roughly characterizes "pre-genomic" theoretical population genetics; contemporary population genetics is considerably more diverse theoretically, especially after the introduction of the coalescent (Wakeley, 2008) and the widespread use of phylogenetic methods on a flood of genetic data.

ley, 1981, 1983, 1987, 2007, 2008; Shott, 2008; Stern, 1994, 2008; Wandsnider, 2008). Similar to the attention paid by paleontologists and paleobiologists (Kidwell, 1997; Kowalewski, 1996; Olszewski and West, 1997; Olszewski, 1999, 2004), time averaging has also seen serious work in zooarchaeological and faunal analysis (Broughton and Grayson, 1993; Grayson and Delpech, 1998; Lyman, 2003) given the importance of diversity indices and other summary statistics whose interpretation is greatly affected by assemblage duration.

My efforts to address the issue of equifinality are included in Chapter 2. Written in 2012 and released on Arxiv.org, this article provided the first analysis of the effects of time averaged samples on the diversity statistics and statistical tests of neutrality that were commonly being employed in cultural transmission research within archaeology. I used agent-based simulation to sample the behavior of neutral and non-neutral transmission models under varying degrees of time averaging, and examined the resulting effects on common diversity statistics and neutrality tests to determine whether transmission bias may have affected the class frequencies we measure. My conclusion, which was then echoed by Premo (2014), is that even moderate amounts of temporal aggregation render standard ”tests” for bias and neutrality unable to discriminate effectively between the two. Since the publication of this work, others have also taken the simulation approaches that Premo and I used. These more recent studies explore how time averaging in our data affects the spatial scale of cultural differentiation, and how the apparent rates of change we measure from archaeological samples scale with duration (Miller-Atkins and Premo, 2018; Porčić, 2014; Perreault, 2018). From the cumulative results of these works, it is quite clear that even moderate amounts of time averaging destroy the ability to treat archaeological samples “as if” they were synchronic.

This lesson led to a significant improvement in archaeological modeling of cultural transmission. In 2013, Kandler and Shennan (2013) moved beyond synchronic model predictions and instead demonstrated how it is possible to extract diachronic or “non-equilibrium” predictions about expected change over time from standard models of neutral and biased cultural transmission . Their work takes a diachronic approach to microevolution modeling rather than trying to "correct" a synchronic modeling approach to match the needs of the data. The task of building on their foundational research will be vital to future success for those exploring cultural phenomena at archaeological scales.

The work of Kandler and Shennan (2013) has been followed by a critique of, and replacement for, the way that archaeologists had been approaching the "model selection" step in the above conceptual approach. Crema (2014) as well as Kandler and colleagues (e.g., Kandler and Shennan, 2015; Wilder and Kandler, 2015; Kandler and Powell, 2018; Kandler and Crema, 2019) have advocated for a "generative approach" to the study of cultural transmission in which model selection is performed against empirical data. The generative approach combines approximate Bayesian model selection (Sisson et al., 2018) with simulation modeling to produce predictive data sets. Based in Bayesian methods, the power of this approach comes from one's ability to estimate the "posterior distribution" of the statistical behaviors one can expect to see from each of a number of transmission models, along with an estimate of how likely each combination of observable statistics would be given the expectations of specific models. This approach allows one to rank statistics derived from archaeological data (e.g., a diversity measure, or the slope of a frequency distribution) by their likelihood to have arisen under each model. One can then examine the likelihoods presented by each model and determine whether there is a single model which could account for the observed data, or–more likely–whether there are still multiple models which could have generated the observed data. Even more importantly, this combination of simulation and model selection allows the study of scenarios with non-stationary parameters, including growing or shrinking populations, and the incorporation of significant population structure in our models (Kandler and Powell, 2018; Rorabaugh, 2014).

This kind of model selection approach, which uses simulation from models to determine the likelihood of observed data under each model, is increasingly common across the sciences and occurs in a number of variants, from parametric and non-parametric bootstrapping (Efron, 1981; Efron and Tibshirani, 1993), multiple model comparisons using a variety of information criteria (Burnham and Anderson, 2002), posterior predictive simulation in Bayesian approaches (Gelman et al., 2013, 1996; McElreath, 2020; Robert, 1994), and approximate Bayesian computation when the likelihood function cannot be evaluated or even formulated in a closed-form equation (Beaumont et al., 2002; Toni et al., 2009; Beaumont, 2010; Csilléry et al., 2010; Marin et al., 2012; Sisson et al., 2018). These kinds of model selection approaches have demonstrated their value in evolutionary biology (see the excellent review by Brown and Thomson, 2018), although as Brown and Thomson note, such techniques are not yet standard practice even given the mathematical sophistication of molecular phylogenetics and other evolutionary subfields. Simulation-based model fitting should be widely applicable in archaeology. In a particularly clear and sophisticated example, DiNapoli and colleagues (2019) combined information-theoretic criteria and simulation from Poisson point-process models to explain the spatial pattern of *ahu* on Rapa Nui, finding that their distribution is most strongly related to the distribution of sources of fresh water. Since approaches like these allow us to quantify sources of uncertainty in our models and judge where models fit and also fail to match our data, they should become standard practice.

1.3 Are There Structural Equifinalities We Cannot "Correct"?

Even with the power of generative modeling and simulation-based model selection, it has proven difficult to distinguish between neutral and biased models of transmission (Kandler and Crema, 2019). This difficulty arises from several sources of structural equifinality that make the archaeological fitting of detailed cultural transmission models to individual populations a difficult or even impossilbe enterprise in most circumstances. One source of equifinality arises from the complex mixture of imitation, teaching, and mixtures of social and individual learning processes that we call "transmission" in real populations (Wimsatt, 2019). Real populations of humans and social animals bear little resemblance to the pure populations of most models employed in the literature today. Another source of equifinality arise from the stochastic nature of the models we necessarily employ and our inability to sample multiple realizations of a stochastic process when we try to fit individual assemblages to transmission models. Any single data point may be compatible with a wide variety of models; only with multiple samples from the same realization of a process can we hope to do model selection with validity and statistical power.

The first source of structural equifinality is a consequence of the basic features of the phenomena we study: real human populations interact in complex ways. They never follow any single "mode" or strategy for adopting cultural information and learning from their peers. Realistic populations always include variation among individuals, and individuals often vary over time in the degree to which they vary the ways they learn or adopt behaviors . Individuals might follow conformist strategies at one point or novelty-seeking tendencies at another. Individuals vary the learning strategies they employ depending upon the type of situation faced, or the kind of trait involved. Thus, we should expect that populations will always be mixtures of cultural transmission modes and learning strategies, and one would expect the statistical signatures of these strategies to present complex statistical profiles. In the worst cases, the contributions of different learning strategies and biases may even "cancel out" at the population level, entirely eliminating our ability to distinguish one model from another. Given the effect of population mixtures, much of the equifinality we encounter in the microevolutionary approach is structural, and will not be resolved through better analytical methods—it is built into the phenomena themselves.

In 2016, I wrote Chapter 3 to explore this issue. This chapter examines the degree to which we can distinguish mixed populations using coarse grained, population level data. I use a variation on the generative approach described earlier. The study pairs agent-based simulations of differing population mixtures and compares them to each other and to a population of unbiased copiers. From each model in a pair, 23 different summary statistics are collected, all of which have been in use in the cultural transmission modeling literature. The simulations incorporates the effects of time averaging and sample size, to determine the interaction between these critical empirical factors and our ability to cleanly separate models from summary data. It accomplishes this task by generating several different sets of predicted data from each simulation pair with differing sample sizes and amounts of time averaging. I use a gradient boosting classifier to determine the degree to which any combination of summary statistics are able to distinguish between models in a comparison, and which observable variables are important for separating and identifying the models (Natekin and Knoll, 2013; Friedman et al., 2000; Hastie et al., 2009).

In the chapter, I conclude that that equifinality is rife among mixture models. While the ability to census an entire population under conditions that lack time averaging permits model discrimination, sampling and time averaging quickly makes these discriminations statistically impossible. For example, it is typically impossible to distinguish a mixture of anti-conformists and conformists from a pure population of unbiased copiers. The effects can simply "cancel out" at the level of the population. Only under the simplest conditions in which observations are synchronic and populations are fully censused is there enough departure from the expectations of neutrality in a population that the classifier can find combinations of predictors that separate the distributions. But when one is limited to using finite samples and/or when samples represent significant intervals of time, it is difficult or impossible to tell which mixture of models may be represented in empirical data. Although the relationship between assemblage sample sizes and population sizes are typically not directly studied in most modeling exercises, our samples of the archaeological record are always samples of artifact discard and deposition from portions of a past population. This inherent sampling issue and thus our inability to discriminate among mixtures of transmission modes seems structural and unavoidable.

It is not hard to understand why our simple models of transmission produce so many avenues for equifinality. Our models of transmission are stochastic and incorporate chance in the processes of learning between individuals, and in the ways in which we model innovation. Chance is a key component of all historical phenomena. While the general claims of Billiard and Alvergne (2018) about the lack consideration of this fact among archaeologists are accurate, most archaeologists accept that stochastic models are essential for modeling complex social behavior, and chance plays an important role in explaining any historical or evolutionary phenomenon.

That said, the structure itself of our models of biased and unbiased transmission (especially for discrete traits) contribute strongly to the potential for equifinality. At their core, each of the models we have attempted to fit to archaeological data are Markov chains that model trajectories of change in integer partitions (Crane et al., 2016). Our models are, structurally, all variations of sampling schemes from distributions within the Poisson-Dirichlet family; when the samples represent unordered partitions, the famous Ewens Sampling formula or distribution results. The Ewens distribution does reflect the underlying probability model for the "infinite-alleles" model of neutral drift (Ewens, 1972), but there is strong convergence in distribution for other models as well (Huillet, 2007). The Ewens distribution can represent the distribution of allelic partitions under selection as readily as it can neutrality (Gillespie, 1977; Grote et al., 2002; Khromov et al., 2018; Sawyer and Hartl, 1985).

As a consequence of sharing this basic structure, it can be difficult to determine if a sample of data derives from any specific member of the family, especially at small sample sizes, and if the number of elements in a partition (artifact classes or categories, in our case) is small. The small departures from a power law distribution, for example, that might be diagnostic in the context of a population census and using many classes (e.g., baby names or dog breed frequencies) are difficult to detect with small samples and using small numbers of classes. The larger the number of partitions (or classes) represented and the larger the sample size used, the larger the number of states that can be empirically distinguished. Relatively speaking, small numbers of classes and small sample sizes lead to small numbers of distinguishable states. The problem we face, therefore, is that our microevolutionary models strongly overlap in their distributions of distinguishable states. The only potential detectable differences are slight variations in how probable any given state is from one model versus another, not its presence or absence in the solution set. Given this, models with small state spaces will be very hard to distinguish from one another.

This fact directly explains how the equifinality between cultural transmission models arises in the discrete case, especially when we only observe a single data point or realization. Determining the best model fit, one typically needs multiple samples from the population under study so that one can quantitatively assess which model is the most likely fit. This step means some of the early studies that focused on the degree to which a single assemblage or several components from a single site compared to expectations of transmission models tended to have poorly defined results. While Kandler's (2013) diachronic, non-equilibrium approach increases the statistical power of results by looking at the likelihood of trajectories of observations rather than single data points, this approach truly becomes powerful in cases in which one can sample enough points through time and across space to manage complications of sample size and time averaging issues.

Ulimtately, the combination of generative methods combined with attention to the size of the modeling state space offers the only practical solution. Our models need to have richer state spaces, so that their predictions are not so strongly overlapping. We need to develop predictions that encompass the quantitative aspects of information flow at more than just a couple of points in time and at more than one location. Richer predictions need to be matched by a richer set of observables that go beyond simple frequency arrays. We need to develop structures that can represent diachronic change and spatial variation, as well as that vary in enough ways to be statistical distinguishable using typical archaeological sample sizes. Only by addressing these needs will we be able to get beyond structural equfinality in our modeling, and distinguish between hypotheses about evolutionary history in the archaeological record.

1.4 Seriation and the Mesoscopic Approach to Cultural Transmission Modeling

In 1995, Carl Lipo and I (Hunt et al., 1995; Lipo et al., 1995, 1997; Lipo and Madsen, 2001, 1997; Lipo, 2001a,b) began systematically exploring seriations as observable units for fitting transmission models to archaeological data. We engaged in this exploration due to the recognition by Dunnell (1970) that seriation automatically incorporates the diachronic nature of our data and includes finite durations for each of the assemblages that make up the seriation. In some of our early work on the subject (e.g., Lipo et al., 1997), we introduced an iterative method for finding deterministic solutions to the seriation problem, by partitioning the full set of assemblages into subsets, where each subset fully meets the requirements for unimodality.

Rather than using seriation in its traditional format to build single linear orderings for all assemblages, this work involved creating multiple subsolutions. Creating multiple solutions accomplishes two important results. First, it ensures that each subset meets the seriation criterion being used (e.g., unimodality or occurrence). Second, the creation of multiple solution directly incorporates the spatial variation present in the history of artifact classes across a region. This latter factor takes advantage of one of the key reasons that classical seriations used a "same local area" criterion to limit the amount of spatial variation one put into a seriation. The spatial restriction is due to the fact that groups of assemblages in different places will produce different orders, and it is impossible to accommodate these different histories using a single linear order.

One way to account for the effect of space on the composition of class frequencies is simply to break the set of assemblages being ordered into the largest groups of different solutions. Each of these valid seriation solution tells one something unique about the history of artifact assemblages in specific places and times. In doing this kind of multiple solution technique, individual assemblages must be allowed to be included into solutions for multiple subsets. By examining how assemblages can fit into multiple seriation solutions provides a way to map how information may be differentially flowing in and out of localities within a region and through time. In practice, a small subset of assemblages do tend to occur in multiple subsets, as Lipo (2001b) found in his seriations of Mississippian ceramics from the central and lower Mississippi River valley.

Lipo also extended deterministic frequency seriation by calculating bootstrap confidence intervals around class frequencies, a step that allows one to assess the unimodality criterion for a set of assemblages while taking into account the likely effects of sampling error. This allows one to: (*a*) determine when two possible orderings for a set of assemblages cannot be distinguished, and thus certain assemblages are "contemporaneous" given the data we have, and (*b*) understand when small deviations from full unimodality are likely just vagaries of sampling error. Much of this early work was still accomplished by manual assortment of assemblages and then software-based confirmation of the significance of candidate solutions via bootstrap analysis (using a Microsoft Excel macro package which still gets download requests today.[2]

More recently, Lipo and I (2015), finishing work begun with Dunnell before his passing, automated the process of finding multiple-subset seriation solutions, by converting the problem to one of graph or tree construction. In that work, which is not included in this dissertation, we outlined an

---

[2] `http://www.evobeach.com/p/seriation.html`

approach to seriation graph construction that employs the bootstrap confidence interval testing and employs heuristics to quickly prune the set of possible solutions. This approach helps avoid, but cannot prevent, the combinatorial explosion that occurs as the number of assemblages increases. In a short research paper, included here as Chapter 4, I examine how employing multiple solution groups affects the size of the "solution space" for the seriation problem and indeed increases the number of possible solutions for a given set of assemblages by orders of magnitude over the permutations available in a straight linear order.

Increasing the number of possible solutions might sound like a bad outcome, and without good heuristics on finding possible linkages within the candidate solutions it would be. But in practice, our "iterative deterministic seriation solution" (IDSS) method proved relatively tractable with around a dozen assemblages, especially with the significant increases in computing power now available to researchers. But it is well worth looking at how to efficiently find solutions when we have 20, 30, or 50 assemblages. Why? Because the larger the set of assemblages we can include, with their variation in artifact class frequencies, the more data we are sampling from the single realization which was the actual history of cultural information in some span of time in a given region. Additionally, larger seriations provide a larger state space within which structural variation can help us distinguish between hypotheses.

In order to improve our efficiency in finding complex solution graphs, I worked with Lipo to examine alternate criteria for forming solutions. There is nothing special about unimodality in a seriation. As Neiman (1990, 1995) has documented, realizations of an unbiased transmission model like Wright-Fisher easily give rise to unimodal rises in the "popularity" of some trait, peak, and then decline, just as Nelson (1916) and Wissler (1916) described in the earliest recognizable "frequency seriations." But transmission also gives rise to other types of patterns and can easily give rise to multimodal distributions for a trait over time as well (some things come back into "fashion", as we well know from contemporary life). Based on this fact, it becomes clear that unimodality is not central to frequency seriation because it is the only pattern possible for transmission among a population of indivduals over time, but because it is a distinctive pattern among many that can be used to find the unique history of information sharing between and among communities. Recognizing this fact provides an opportunity to look for other ways of doing frequency seriation that yield equivalent results, but are more general and more efficient.

Chapter 5 describes our work comparing unimodality to other possible ordering algorithms, and in particular distance-minimization, building upon Kadane's (1971) earlier work. The principle we use in our reconceptualization of the seriation method is to find seriation solution graphs that globally minimize the total amount of change between neighboring pairs of assemblages. Because this method alludes to the ideas of mathematical "smoothness" and "continuity," we dubbed the method "continuity seriation." With continuity seriation, the efficiency of the calculation dramatically increases the size of data sets that can be analyzed, by providing a roughly 25x speedup in evaluating solutions (conservatively estimated). The expansion of the possible size of the solution space is a good thing when we think about cultural transmission models and their possible outcomes over many assemblages that span a region. If we treat a seriation solution for a set of assemblages as a realization of cultural transmission outcomes within a region over time, it is likely that some models of regional transmission are compatible with that realization (seriation) and that many will not be. By expanding the size of the potential solution space, we reduce the potential for equifinality.

With continuity seriation available as a technique for obtaining solutions for larger numbers of assemblages at once, it becomes possible to tackle the question of whether seriations might be diagnostic of particular classes of evolutionary histories at the mesoscopic scale of analysis. I examine this question in Chapter 6, by introducing a method of formalizing hypotheses about the regional structure of cultural transmission. A transmission scenario is defined as a candidate regional history of the social network between communities and how it might change over time. The results of repeated acts of cultural transmission over this evolving social network results in differential adoption and persistence of stylistic or neutral- behaving artifact classes (*sensu* Dunnell 1978) across the whole set. Standard social network models are typically synchronic snapshots, so the formalism adopted here is that of an "interval temporal network" to model how connectivity changes over time (Holme and Saramäki, 2012).

In Chapter 6, I examine four different regional scenarios. These scenarios include configurations for: (*a*) "complete" networks, where communities existing at any particular time index are all in communication; (*b*) A "nearest neighbor" network in a long, thin configuration, where neighboring communities are connected to spatial neighbors along a river, with a few longer-distance connections (perhaps exogamous marriage relationships or trading relationships); (*c*) A "nearest neighbor" network in a compact, square configuration, where neighboring communities connect to each other in various directions, with a small number of longer-distance connections, and (*d*) A scenario in which a single regional population with complete network clusters of neighbors and a few longer-distance links between clusters splits partway through the time course of simulated evolution to form distinct "lineages" which are no longer in communication.

The goal of this study is to outline a potential method for: (*a*) determining whether sets of transmission scenarios, expressed as interval temporal network models, can be distinguished in theory (i.e., are they equifinal), and (*b*) can we use the statistical models we create in exploring the equifinality question to assess the "fit" of empirical datasets to our transmission scenarios? To that end, I construct a simulation framework for sampling cultural trait frequencies across regional transmission scenarios, while applying realistic sampling and time averaging to the sampled data. Our IDSS seriation method with continuity criterion then produces simulated seriations from replicated simulations from each regional transmission history. I then consider how to best represent the structure and topology of seriation solutions in a statistical model. In this work I employ the eigenvalue spectrum of the Laplacian matrix for the seriation graph, since the Laplacian is known to capture most of the topological information present in many classes of graphs. A machine learning classifer model, trained on 90% of the simulated seriation data, was employed to predict the most likely data generating process (transmission scenario) for a hold-out test set comprising the remaining 10% of simulated seriations.

In general, the results are promising. In the test set results, we can tell the difference between lineage splitting, nearest neighbor, and more homogeneous social networks. The results indicate that we cannot distinguish between the two spatial configurations of nearest-neighbor scenarios, at least with the size of seriation solutions employed and with the Laplacian spectra as predictor variables. This tells us something about the scale and level at which we can hope to frame mesoscopic models of cultural transmission. Finally, I examined seriation results from the Late Prehistoric ceramic data in the central Mississippi River valley using a trained gradient boosting classifier. The results from this analysis appears that the results are consistent with a lineage splitting event. Overall, these results are consistent with previous conclusions reached using overall archaeological evidence for this data set (Lipo, 2001b). This kind of generative approach with seriation graphs as the unit of observation has considerable promise, and this initial work is simply a down payment on exploring issues such as the sample sizes needed to resolve different classes of regional scenarios.

1.5 Dependency Graphs and Incorporating Structured Information In Cultural Transmission Studies

Much of this dissertation consists of my efforts to understand the flow of cultural information through space and time. One area that remains unexplored are the methods needed to understand how the *content* of culturally transmitted information changes over time, and how the kind of information may affect its transmission. Simple cultural transmission models tend to treat cultural traits in the manner of "bean bag genetics" – as markers which come and go and are subject to innovation but have little structure among themselves. Cultural information and the skills that people inherit and pass on with that information are nothing like simple markers. Wimsatt (Wimsatt and Griesemer, 2007; Wimsatt, 2013, 2014, 2019) has made this the central focus of his work on cultural evolution, and has brought together researchers with a variety of disciplinary foci to make development central to the study of cultural transmission. Within archaeology, Mesoudi and O'Brien (2008b) explored structured relations between cultural traits. Tostevin (2019) has richly developed this idea and has combined work on trait structure with Wimsatt's idea that some cultural traits provide "scaffolding" needed to learn others. He argues for the creation of "thick descriptions" that include the details of how social learning and cultural transmission articulate with the actual physical processes involved with technology. For example, he explores the relations between the physics of flintknapping and the processes of learning to flintknap and demonstrates that we can articulate the actual physics of the technology with the homologies we see over longer spans of time as methods are taught and learned.

The fruitful marriage of learning theory, the details of specific technologies, and longer-term patterns in transmission is an exciting development in evolutionary archaeology. This avenue of research represents early steps in moving beyond simple models to explanatory models that use the cultural transmission framework to answer real questions about our evolutionary history. Successfully achieving this goal requires us to develop new tools required to make the articulation. These tools include the establishment of meaningful observable units and the determination of the statistical properties of those observables.

The final chapter in this dissertation consists of my attempt to address these needs. I wrote this chapter in 2015 for a volume on social learning in Neandertals and early modern humans. This volume focused on a "learning hypothesis" for behavioral modernity in the Upper Paleolithic (Nishiaki et al., 2013). In Chapter 7, I examine a case of "structured information" in which traits are modeled as having prerequisites. This situation often occurs in learning related to technology. For example, the acquisition of some skills may not be possible until we have mastered other skills. Conceptually, we can represent the relations among cultural traits or artifact classes using dependency trees. In these trees, nodes that are represented as higher in the tree are prerequisites for traits that are lower down. Using this model of dependencies, I modeled how different learning models such as individual trial and error and targeted teaching by a peer produced cultural repertoires of different structure and richness. Like elsewhere in my dissertation, I employed a simulation approach and examined the "knowledge graphs" that simulated individuals have after many rounds of transmission while also conditioning each simulation run with different rates of teaching and individual innovation. Since the dependency structures and traits are abstract, the variables are trees of traits. I then analyzed the topology and symmetries of the trees of traits to determine their structure, using tools from algebraic graph theory (Godsil and Royle, 2001). These structures varied from those that were deep and broad to those that were shallow and narrow. The results of my study supports the hypothesis that the evolution of mechanisms of teaching and apprenticeship would lead to enriched cultural repertoires and growth in cultural diversity.

Within the context of this dissertation, this study is significant because it demonstrates, once again, that studying cultural transmission within archaeology requires careful consideration of how we structure our observations. As Dunnell (1971, 1986) repeatedly pointed out, the observational units we employ are not a given. The artifact class frequencies we construct for one purpose do not necessarily mean they are informative for another purpose. We must build observational units using a combination of good old fashioned artifact classification as well as all of the mathematical, statistical, and machine learning sophistication we can muster. Only through this combination of efforts can we fruitfully use cultural transmission models in archaeology to tackle questions of evolutionary history. And we must tackle questions of evolutionary history if we are going to avail ourselves of the bodies of theoretical machinery, from evolutionary game theory to decision-theoretic modeling, to form complete evolutionary explanations (e.g., Gintis, 2014).

CHAPTER 2

Neutral Cultural Transmission in Time Averaged Archaeological Assemblages

Abstract Neutral models are foundational in the archaeological study of cultural transmission. Applications have assumed that archaeological data represent synchronic samples, despite the accretional nature of the archaeological record. Using numerical simulations, I document the circumstances under which time-averaging alters the distribution of model predictions. Richness is inflated in long-duration assemblages, and evenness is "flattened" compared to unaveraged samples. Tests of neutrality, employed to differentiate between biased and unbiased models, suffer serious problems with Type I error under time-averaging. Estimation of population-level innovation rates, which feature in many archaeological applications, are biased even without time averaging, but have sharply increased bias given longer assemblage durations. Finally, the time scale over which time averaging alters predictions is determined by the mean trait lifetime, providing a way to evaluate the impact of these effects upon archaeological samples.

Source Posted to Arxiv.org (`https://arxiv.org/abs/1204.2043`)

2.1 Introduction

The evolutionary study of culture today crosses many disciplines and employs a variety of experimental and observational methods to study its subject matter. What makes the archaeological record unique as a source of data concerning the evolution of culture is time depth, creating the possibility of studying both the unique histories of human groups and the evolutionary processes that shape those histories. Archaeology is not unique in studying temporal data on human activity, but like our colleagues in paleobiology, we study an empirical record that is unlike the time-series data available to disciplines such as economics or epidemiology (e.g., Arrow, 2009; Keeling, 2005; Keeling and Rohani, 2007; Kendall and Hill, 1953; Rothman et al., 2008). The archaeological record is not a sample of measurements from individual moments in time stacked together into a sequence. Instead, archaeological deposits are almost always accretional palimpsests, representing cumulative artifact discard over durations of varying length (Bailey, 2007, 1981; Binford, 1981; Schiffer, 1987; Stein, 1987). Thus, when archaeologists count the richness of faunal taxa in an assemblage, or measure the relative frequencies of ceramic types, the data obtained summarize the bulk properties of artifact discard and deposition over significant spans of time, often with nonconstant rates of accumulation.[1] We refer to assemblages which are accretional in this manner as "time averaged."

A growing number of studies apply cultural transmission models to artifact assemblages by comparing the predictions such models make for the richness, diversity, or frequency distribution of cultural traits, to counts or frequencies of artifact classes (e.g., Bentley and Shennan, 2003; Bettinger and Eerkens, 1999; Eerkens and Lipo, 2007; Jordan and Shennan, 2003; Lipo and Madsen, 2001; Perreault and Brantingham, 2011; Premo and Scholnick, 2011; Scholnick, 2010; Shennan and Wilkinson, 2001; Shennan, 2011; Steele et al., 2010). The question is, are model predictions comparable to archaeological measurements? Given the time averaged structure of most archaeological

---

[1] As well as the action of various post-depositional and taphonomic processes, of course.

deposits, I suspect the answer is no. Transmission models developed outside archaeology are typically constructed to make predictions concerning variables observed at a point in time. To date, almost none of the archaeological literature employing cultural transmission models has taken this "time averaging" effect into account and modified the way predictions are made to match the nature of the phenomena we measure (cf. Bentley et al., 2004). Evaluating the effects of temporal aggregation upon the predictions made by cultural transmission models is the first step in understanding how to rewrite and adapt transmission models to understand their dynamics given time averaged observations.

In his dissertation, Neiman (1990) considered a potential source of time averaging effects in diachronic assemblages: variation in discard rates across traits. With respect to this particular effect within accretional deposits, Neiman's results suggested that the predictions made by a neutral model of cultural transmission were directly applicable to the relative frequencies of traits as we would measure them in a time averaged assemblage. Nevertheless, there is good reason to consider the effects of aggregation directly, outside of variation in discard rates. Paleobiologists, for example, have documented systematic differences between living and fossil assemblages, including increased species richness, reduced spatiotemporal variance in taxonomic composition, and flatter species abundance curves in time averaged assemblages (Olszewski, 2011; Tomašových and Kidwell, 2010a,b). Lyman (2003) extended these results to zooarchaeology, noting that time averaging can be a significant problem when the process one is applying or studying occurs over a shorter time scale than the empirical record available to study its properties (see also Grayson and Delpech, 1998). This relation between time scales is applicable to cultural transmission modeling as well.

Archaeologists now employ a variety of cultural transmission models, which differ in the kind of variation and traits they describe and the copying rules and evolutionary processes they incorporate. Discrete models describe individual variants or traits by their count or frequency in a population and are foundational for the study of stylistic variation in many artifact categories (e.g., pottery). The simplest discrete model is random copying in a well-mixed population with innovation, representing neutral variation with the stochastic effects of drift. We frequently construct more complex models of transmission bias by adding additional terms or frequency-dependent copying rates to the basic unbiased copying model (Cavalli-Sforza and Feldman, 1973a,b, 1981; Boyd and Richerson, 1985). Thus, an understanding of the effects of time averaging upon neutral transmission will be informative about many (if not all) of the discrete transmission models in use by archaeologists today, and forms the focus of the present study.

I report the results of numerical simulations designed to observe neutral transmission using variables employed in the archaeological literature, aggregated over time at a variety of intervals designed to mimic a wide range of "assemblage durations." In Section 2.2 I describe the relationship between neutrality, unbiased copying, and the separate but related concept of "drift," followed by a review of the quantitative properties of the well-mixed neutral Wright-Fisher infinite-alleles model in Section 2.3. Section 2.4 outlines the simulation model employed to study time averaging in this paper, including model verification and testing, and the algorithm used to effect temporal aggregation within the simulations. Section 2.5 presents the results of simulating unbiased cultural transmission for a variety of innovation rates and assemblage durations, and Section 2.6 summarizes the effects seen and points to next steps in reformulating our cultural transmission models for archaeological contexts.

2.2 Conceptual Structure of Neutral Cultural Transmission

In his classic article "Style and Function: A Fundamental Dichotomy," Dunnell (1978) proposed that many aspects of an artifact would play little or no role in its engineering performance, and thus have no impact on the fitness of individuals employing it. In other words, some attributes of artifacts are neutral with respect to selection. This has been widely misinterpreted as a claim that the artifacts themselves are neutral or have no fitness value, which is not the case. Dunnell was saying that if one describes an artifact solely using attributes which have equal cost or performance, the resulting classes meet the definition of neutral variation.

Fraser Neiman (1990) first connected Dunnell's identification of style as selectively neutral variation, to population genetic models designed to describe genetic drift. His dissertation considers a wide range of cultural transmission models, especially those described by Cavalli-Sforza and Feldman (1973a,b, 1981) and Boyd and Richerson (1985). Neiman employed simulation to calculate the consequences of both individual processes as well as processes combined with various archaeological factors such as variable rates of artifact discard. In this work, Neiman pioneered virtually every technique used by archaeologists today to model and study cultural transmission. The discipline as a whole was introduced to this work in his now classic 1995 article (Neiman, 1995), in which the dynamics of Woodland ceramic variation were explicitly modeled as a random copying process.

Despite the fact that there are multiple ways that neutrality can arise as a population level effect, there is a tendency today to equate neutrality with "drift" in the archaeological literature on cultural transmission. For example, Bentley et al. (2004, p.1443) offer a fairly typical description of unbiased cultural transmission as "random genetic drift, which describes how the diversity of variants evolve when the dominant process is one of random copying." In fact, drift and the copying rules that create population-level trait distributions are different and independent aspects of a transmission system. Before we turn to the details of a formal model for unbiased, neutral transmission, it is worth reviewing the conceptual elements that make up such models.

Drift is a feature of any stochastic transmission model in a finite population, regardless of whether selection or bias is also present in the model. Sewall Wright gave the name "genetic drift" to the random fluctuations in gene frequency that occurred because some individuals might be the source of many genes in the next generation, and others none at all. Translated into a cultural model, drift occurs when some individuals, by random chance, are imitated or copied and others are not. In an infinite population, by contrast, the variants held by individuals would be sampled at their exact frequencies in the population, and thus there would be no stochastic "wiggle" in trait frequencies. This is reflected in population genetics by the famous "Hardy-Weinberg" equilibrium, where in the absence of selection or other forces, gene frequencies stay the same from generation to generation. This means that we can easily have *neutrality without drift*, in an infinite population. In a large but still finite population, we can expect drift to have very tiny, potentially even unmeasurable effects upon the trajectory of trait frequencies.

Drift, moreover, occurs in combination with a variety of inheritance rules, mutation models, and in combination with natural selection. In small populations, we can expect drift to be a factor when examining the engineering properties of ceramics and the relative fitness of firing technologies, or the fitness of foraging strategies. Whenever such traits are learned and passed on within small, finite populations, the stochastic aspect of who learns from whom will create fluctuations in variant frequencies that have nothing to do with the performance or survival value of traits, or the prestige of those we choose to learn from or imitate. In other words, we can have *drift without neutrality*. In small enough populations or during bottlenecks, even adaptive technologies and knowledge can be lost to drift (Henrich and Boyd, 2004; Henrich, 2006). We should always be on the lookout for the effects of drift, especially as population sizes get smaller as we go back in time. Drift is not a model of human social learning; it is a consequence of finite populations, injecting stochastic noise into the dynamics of a system that affects our ability to cleanly fit models and test hypotheses.

Neutrality, by contrast, is a population level phenomenon, arising when there is no net force systematically favoring certain variants over others for a particular dimension of variation. Most commonly, of course, we mean that there is no natural selection that favors some alleles over others, but from a mathematical perspective, the transmission bias rules of Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985) are equivalent to selection models.[2] The simplest way for neutrality to arise is for individual social learning to be "unbiased." Unbiased transmission models always yield population-level neutrality for the traits being passed, because the probability of imitating any specific trait is simply proportional to its frequency among individuals in the population. The Wright-Fisher model is one of the earliest stochastic models in population genetics (Provine, 1989, 2001; Wright, 1931), and was originally created to describe the process of genetic drift and its effects in combination with other evolutionary processes. Following Kimura's theory of neutral alleles, Wright-Fisher is also used to describe the evolution of populations in which variants are selectively neutral. Elaborations of the basic Wright-Fisher model add mutation, selection, loci with multiple alleles, and multiple loci with interactions between loci (see esp. Crow and Kimura, 1970; Ewens, 2004).[3]

But unbiased copying is not the only source of neutrality among variants, and it is important to keep this in mind when selecting models to test as explanations for archaeological phenomena. In any realistic human population, there will be heterogeneity in social learning rules, with individuals using different rules for different traits, or kinds of traits, and perhaps having individual propensities for conformism (all other things being equal) or pro-novelty bias (Mesoudi and Lycett, 2009). A population which is heterogeneous for such rules may display the characteristic frequency distributions of conformity or pro-novelty biased if we are able to observe small numbers of transmission

---

[2] In this paper I leave aside the relationship between "natural" and "cultural" selection, and transmission biases, since such issues are largely philosophical and theoretical and do not affect the nature of the models we employ for quantitative analysis of cultural variation.

[3] And, the Moran family of models mirrors the Wright-Fisher models, with overlapping generations, by representing dynamics as continuous-time stochastic processes. Moran models are likely the best framework for modeling cultural transmission when the exact temporal dynamics matters. In this paper I follow archaeological convention by employing the more familiar Wright-Fisher discrete generation framework.

events or individual transmission chains, while simultaneously cancelling each other out at the level of the population. In other words, heterogeneity is a major source of equifinality between different models of social learning, when observed through population-level trait frequencies. No archaeological applications of cultural transmission models today have employed heterogeneous models, probably because the theory behind such models is not well-studied. But this is clearly a frontier for future research since homogenous models poorly reflect what occurs in real human populations.

Returning to unbiased models of transmission, we face a further choice in selecting a specific model to employ or study. In addition to the copying rules, we must specify an innovation rule. Such a rule answers questions like: how do new variants enter the population, can variants be invented multiple times independently, and is there a constrained range of variation for a particular dimension of an artifact? For example, painted design elements on a ceramic pot offer a "design space" of possibilities that is potentially unbounded, even if only a tiny fraction of possible designs occur in any archaeological context. Such attributes are best modeled by the "infinite allleles" innovation model. In contrast, stylistic aspects of lithic tools may be sharply constrained by the technology and materials themselves, and may be best modeled by innovation among a small set of variants, with the material constraints causing frequent "reinvention" of the same shapes over and over. Such attributes are best modeled by constraining the design space, and employing a finite or "k-alleles" version of the unbiased model. Since Neiman's pioneering work, most archaeological applications of neutral models have employed the "infinite alleles" variant of the Wright Fisher model (WF-IA)(Kimura and Crow, 1964). Therefore, in the remainder of this paper, I focus on the unbounded model of neutral evolution with innovation, since it is relevant to a large number of archaeological contexts and artifact categories, but the reader should be aware that the models with a constrained number of variants may be hugely important in specific archaeological contexts, and are underexplored in the archaeological literature.

2.3 Unbiased Transmission: The Wright-Fisher Infinite-Alleles Model

WF-IA is a stochastic process that models unbiased transmission within a fixed-size population as multinomial sampling with replacement, with a mutation process that adds new variants to the population at a known rate. After describing the model, I review the sampling theory of Ewens (1972), which gives the distribution of variants expected in small samples taken from the population as a whole. The sampling theory, rather than the distribution of variants in the full population, is both well-understood, and most relevant to archaeologists, who are always sampling an empirical record of past artifact variation.

The well-mixed neutral Wright-Fisher infinite-alleles model (Kimura and Crow, 1964) considers a single dimension of variation ("locus") at which an unlimited number of variants ("alleles") can occur, in a population of N individuals.[4] The state of the population in any generation is given in several ways: a vector representing the trait possessed by each individual (census), a vector giving the abundance of each trait in the population (occupation numbers), or by the number of traits represented in a population by a specific count (spectrum).

In each generation, each of N individuals selects an individual at random in the population (without respect to spatial or social structure, hence "well-mixed"), and adopts the trait that individual possessed in the previous generation.[5] Equivalently, a new set of N individuals are formed by sampling the previous generation with replacement. At rate μ for each individual, a new variant is added to the population instead of copying a random individual, leading to a population rate of innovations $\theta = 2N\mu$ (Ewens, 2004), with no "back-mutation" to existing traits.[6] An important consequence of

---

[4] Conventionally, the model treats a diploid population, in which N individuals each have two chromosomes and thus there are always 2N genes tracked in the population. The haploid version is more appropriate for modeling cultural phenomena, and thus formulas given in this paper may differ from those given by Ewens (2004) and other sources by a factor of two. For example, the key parameter θ is defined as $2N\mu$ rather than the common genetic definition $4N\mu$.

[5] An individual can select themselves at random since sampling is with replacement, and this would be equivalent to "keeping" one's existing trait for that generation.

[6] It is important to note that θ is not a measure of the "diversity" of traits in the population, as it has been employed

this innovation model is that each variant is eventually lost from the population given enough time, and replaced with new variants. Thus, there is no strict stationary distribution for the Markov chain describing WF-IA, although there is a quasi-stationary equilibrium in which the population displays a characteristic number of variants, with a stable frequency distribution governed by the value of θ (Ewens, 2004; Watterson, 1976).

Beginning with a now-classic paper Ewens (1972) constructed a sampling theory for the neutral WF-IA model, allowing the calculation of expected moments and frequency distributions for small samples (compared to overall population size) (see Ewens, 2004, for a complete summary of results on the sampling theory). In what follows, we assume that a neutral WF-IA process is running within a population of size N. At some moment in time after the population has reached its quasi-stationary equilibrium, we take a sample of n individuals, where the sample is small compared to the population size ($n \ll N$). We then identify the variants held by each individual. The total number of variants seen in the sample will be denoted by k, or k_{obs} depending upon context.

Given such a sample, Ewens (1972) found that the joint distribution of the variant spectrum (a_i represents the number of variants represented i times in a sample), given the population innovation rate (θ), is given by the following formula (now known as the Ewens Sampling Distribution):

$$\mathbb{P}_{\theta,n}(a_i, \ldots, a_n) = \frac{n!}{\theta^{(n)}} \prod_{j=1}^{n} \frac{(\theta/j)^{a_j}}{a_j!} \tag{2.1}$$

where $\theta^{(n)}$ is the Pochhammer symbol or "rising factorial" $\theta(\theta+1)(\theta+2)\cdots(\theta+n-1)$. In most empirical cases, we cannot measure (or do not set through experiment) the value of θ, so a more useful relation is the distribution of individuals across variants (i.e., the occupation numbers), conditional upon the number of variants k_{obs} observed in a sample of size n:

---

in several archaeological studies, but is instead a *rate* parameter of the model.

$$\mathbb{P}(n_1, n_2, \ldots, n_k | k_{obs}) = \frac{n!}{|S_n^k| k! n_1 n_2 \cdots n_k} \tag{2.2}$$

where $|S_n^k|$ denote the *Stirling numbers of the first kind*, which give the number of permutations of n elements into k non-empty subsets (Abramowitz and Stegun, 1965). The latter serves here as the normalization factor, giving us a proper probability distribution.

From Ewens's sampling theory, and in particular Equation 2.2, a number of useful measures can be derived, relevant to archaeological applications. In this study, I focus upon the most commonly used: statistical tests of neutrality, estimation of innovation rates (θ), and the evenness with which variants are represented in the population (as revealed by several diversity measures).

2.3.1 Statistical Tests for Neutrality

Because Equation (2.2) requires no unobservable parameters, it serves as the basis for goodness-of-fit tests between empirical samples and the neutral WF-IA. The two most important such tests are the Ewens-Watterson test using the sample homozygosity and Slatkin's "exact" test (Durrett, 2008; Ewens, 2004; Slatkin, 1994, 1996, 1994, 1996).[7] Both have been adopted for use by archaeologists, beginning with Neiman (1995) and Lipo (2001b), who described Watterson's work in detail, and more recently, applications of Slatkin's exact test by Steele et al. (2010) and Premo and Scholnick (2011).

The Slatkin test makes no assumptions concerning the process underlying an alternative hypothesis to neutrality, whereas the Ewens-Watterson test examines the observed heterozygosity at a locus versus the expected heterozygosity predicted by Ewens sampling theory. Slatkin's test does not employ the concept of heterozygosity, and relies only upon the "shape" of the Ewens Samping Dis-

---

[7] There are several other important tests of neutrality when dealing with DNA sequence data, including Tajima's D, the HKA test, and the McDonald-Kreitman test (Durrett, 2008). Because their assumptions are highly specific to the structure of sequence data, I omit consideration of them here.

tribution given a specific innovation rate. As a result, archaeologists should prefer Slatkin's test for examining the fit of a synchronic sample of variants to the null hypothesis of neutrality. Slatkin's test is modeled upon the Fisher exact test for contingency tables. Where the Fisher exact test determines the probability of an unordered configuration from the hypergeometric distribution, Slatkin's test determines the probability of a sample of traits (characterized by occupation numbers) with respect to Equation 2.2.

There are two methods for determining how probable a given sample is, with respect to the ESD. For relatively small n and k, it is possible to enumerate all possible combinations ($\mathbf{C}$) of the n individuals among k variants. Each configuration ($c_j \in \mathbf{C}$) then has a probability given Equation 2.2, as does the observed configuration (c_{obs}). With larger sample sizes and values of K_{obs}, it becomes impractical or simply time consuming to enumerate all possible configurations and thus determine the likelihood of an observed sample. In such cases, Monte Carlo sampling of configurations from the Ewens Sampling Distribution is used. We then determine the total probability mass of all configurations (enumerated or sampled) whose probability are less than or equal to the observed configuration:

$$\mathbb{P}_e = \sum_{c_j \in \{\mathbf{C} : P(c_j \mid k) \leq P(c_o \mid k)\}} \mathbb{P}(c_j \mid k) \qquad (2.3)$$

$\mathbb{P}_e$ then represents the Fisherian p-value of the sample with respect to the Ewens Sampling Formula, and thus can be interpreted as a test of the hypothesis that the sample was drawn from a neutral dimension of variation which followed the WF-IA copying model. The $\mathbb{P}_e$ value for a given sample gives the tail probability of its occurrence given the ESD. Thus, if we take a sample of size 100 in a population with innovation rate $\theta = 0.1$, and identify two variants with counts 51 and 49, we might not be surprised to see a $\mathbb{P}_e$ value of 0.01181, indicating that such a sample is highly unusual for a WF-IA process. On the other hand, in the same sample of size 100, if we identify four variants, with counts 55, 38, 6, and 1, this seems a much more typical result of an unbiased copying process. Indeed, the $\mathbb{P}_e$ value of 0.48544 confirms that we should expect to see such samples quite often.

2.3.2 Estimation of Innovation Rates

The behavior of the WF-IA neutral model is governed by the innovation rate (θ). Recall that $\theta = 2N\mu$, and thus represents the population-level rate at which new variants enter the population. In general, for low values of the innovation rate ($\theta < 1.0$), the process is "drift-dominated," and one or a small number of variants dominate the population. At innovation rates above 1.0, which implies that every single "generation" incorporates one or more new variants, the process is "mutation-dominated," and more variants are maintained at intermediate frequencies in the population.

Thus, estimation of the innovation rate from empirical data is of great interest when investigating empirical cases. If we measure the number of variants (K_n) in a sample of artifacts of size n, the sampling theory gives the following probability distribution (Ewens, 2004, Eq. 3.84):

$$\mathbb{P}_\theta(K_n = k) = \frac{|S_n^k|\theta^k}{\theta^{(n)}} \tag{2.4}$$

This is a somewhat inconvenient distribution to work with directly, since calculating the Stirling numbers and rising factorials is both analytically difficult and computationally expensive, but the expected value of K_n has a simple form:

$$\mathbb{E}(K_n) = \frac{\theta}{\theta} + \frac{\theta}{\theta + 1} + \frac{\theta}{\theta + 2} + \cdots + \frac{\theta}{\theta + n - 1} \tag{2.5}$$

K_n is the sufficient statistic for θ, containing all of the information required to calculate the maximum likelihood estimate of the innovation rate ($\hat{\theta}$) from an empirical sample. This is done numerically by finding the value of θ that maximizes the likelihood function of Equation 2.4, or equivalently, finding the value of θ for which the expected value of K_n given Equation 2.5 is equal to the observed number of variants in a sample (since the full distribution may not have a closed-form likelihood function). In the archaeological literature, Neiman (1995) introduced this estimator of θ and called it t_e. With larger samples, Watterson (1975) showed that $k \ / \ \log n$ is a good approximation for the MLE estimator (Durrett, 2008).

Despite the fact that this estimator (and its approximations) are the best that can be achieved from samples, Ewens (1972) showed that all such estimates of θ are biased. Simulations demonstrate, furthermore, that $\hat{\theta}$ (or t_e) is an overestimate of the actual value, and that the amount of bias increases with θ itself (Ewens and Gillespie, 1974). In addition, the variance of the estimator is quite large, and decreases very slowly with increased sample size (Durrett, 2008). The situation is quite different using the "infinite sites" model of neutral evolution and DNA sequence data, where there are excellent and nearly unbiased estimators of theta.

But with the WF-IA and no additional structure to "traits" or alleles, it is very difficult to estimate the innovation rate with any accuracy, or determine whether two samples come from populations with the same innovation rate, or different rates. This fact calls into serious question the degree to which t_e is useful in archaeological analysis, either for estimating innovation rates in past populations, or as a measure of richness or diversity across assemblages or samples. These caveats apply to estimates of innovation rates and t_e given synchronic samples; the effects of time averaging on theta estimation have not been previously documented, and are addressed in Section 2.5.3.

2.3.3 Diversity Measures

The amount of variation expected in a sample is an important quantity, given that we would clearly expect transmission models incorporating bias terms to differ from unbiased or neutral models (e.g. Kohler et al., 2004). Conformist transmission should result in smaller numbers of variants than expected under unbiased transmission, and of course anti-conformist, or "pro-novelty", biases should result in larger numbers of variants being maintained, on average. But beyond helping us assess goodness-of-fit to an unbiased copying model, comparing the number of variants in a sample (K_n) either to a model, or between assemblages, is difficult without reliable estimates of the population-level innovation rate (θ). Since this is inherently difficult and inaccurate, we might ask instead what the evenness of variants is across our samples, since both innovation rates and different models of cultural transmission have clear implications for the diversity of traits we observe.

In the archaeological literature on cultural transmission, the most important evenness measure is t_f, which is a summed estimate of dispersion given trait frequencies Neiman (1995):

$$t_f = \frac{1}{\sum_{i=1}^{k} p_i^2} - 1 \tag{2.6}$$

To make this measure easier to compare across different innovation rates, it is convenient to normalize. Wilcox's "index of quantitative variation," does so, and varies between 0 (when all cases belong to a single category), and 1 (when all cases are evenly divided across categories) (Wilcox, 1973):

$$IQV = \left(\frac{k}{k-1}\right)\left(1 - \sum_{i=1}^{k} p_i^2\right) \tag{2.7}$$

Paleobiologists have found that fossil assemblages have considerably "flatter" species diversity curves compared to living communities, and I expect that time averaging will have the effect here of pushing IQV towards 1.0 compared to its value in unaveraged samples.

2.4 Methods

In this research, I employ a "forward-time" approach to computational modeling of unbiased cultural transmission, by contrast to most modeling in theoretical population genetics today, which employs the coalescent or "backward-time" approach (Kingman, 1977; Durrett, 2008; Wakeley, 2008). In archaeological research, we are interested in the entire distribution of variants which transmitted through the population, samples of which may be deposited and become part of the archaeological record regardless of which variants ultimately leave descendants in later generations. Forward-time approaches evolve a population in steps, applying rules for the generation of variation, copying between individuals, innovation, and sometimes population dynamics.[8] Several well-tested forward-time population genetic frameworks exist, including a very flexible framework called **simuPOP** (Peng et al., 2012; Peng and Kimmel, 2005).

In this research, I employ a framework written by the author specifically for cultural transmission simulations. This project calls for integrating computation models of archaeological classification and seriation, which require code beyond that supplied by population genetics frameworks. My simulation codebase is called **TransmissionFramework**, and is available as open-source software.[9] **TransmissionFramework** runs on any platform capable of supporting a Java 1.6+ runtime, with optional scripts requiring Ruby 1.9+.

2.4.1 Model Verification

Simulation modeling plays an increasingly important role in scientific inquiry, to the extent that computational science is now recognized as a third branch of physics, along with the pre-existing

---

[8] Forward-time approaches are not necessarily equivalent to "agent-based models," but ABM techniques are useful in implementing forward-time models.

[9] **TransmissionFramework** can be downloaded or the code examined at `http://github.com/mmadsen/TransmissionFramework`.

theoretical and experimental branches (Landau and Binder, 2005). Indeed, as theory becomes more complex and realistic, we often cannot directly solve theoretical models and derive predictions that should be measurable by experiment. Computational science sits between theory and experiment, allowing us to understand the behavior and dynamics of complex theoretical models, and calculate predictions that can be used for experiment or hypothesis testing.

The problem of assessing simulation model quality is important enough that the Department of Energy and the Air Force Office of Scientific Research requested that the National Research Council study the foundations of verification, validation, and uncertainty quantification (VVUQ) activities for computational models in science and engineering. Their draft report forms the basis of my approach to verification and uncertainty analysis in this research (Committee on Mathematical Foundations of Verification Validation and Uncertainty Quantification, National Research Council, 2012).

Verification answers the question, "how accurately does a computational model solve the underlying equations of a theory for the observable quantities of interest." Given that we know the true value of θ which drives our simulation runs, it is possible to calculate the expected number of variants at stationarity, and use this to verify that **TransmissionFramework** is correctly implementing the WF-IA. The expected number of traits is a good validation estimate because the number of variants present in a sample will be sensitive to the relative rates of copying and innovation events being handed correctly in the simulation code. Errors in handling these events in software will be magnified across many individuals over many simulation steps.

Since θ is known, the mean value of K_n is well approximated by:

$$\mathbb{E}_\theta(K_n) = \int_0^1 (1 - (1-x)^n)\,\frac{\theta}{x}(1-x)^{\theta-1}dx \tag{2.8}$$

Using Equation (2.8), I compared expected K_n to the average of k_n for a large sample of simulation runs. To ensure that behavior is correct across a range of useful θ values, I performed multiple simulation runs at θ values ranging from 2 to 40, for 5000 generations in a simulated population of 2000 individuals. Each parameter combination was represented by 3 simulation runs. The initial transient behavior of the model is discarded from data analysis by skipping the first 750 generations, given the mixing time analysis by Watkins (2010). At each time step in a simulation run, the simulator took a sample of 30 individuals and tabulated the traits held by those individuals, and recorded the value of K_n. This yielded 408,478 samples across validation runs.

Theta	$\mathbb{E}(K_n)$	Simulated $\bar{K}_n$	Sim. Stdev K_n
2	6.054	6.511	1.838
4	9.022	8.991	2.269
8	12.869	12.616	2.464
12	15.397	15.306	2.571
16	17.228	17.187	2.569
20	18.629	18.737	2.486
40	22.601	22.693	2.253

Table 2.1: Comparison of expected K_n from (2.5) with simulated values from WF-IA model, for θ values from 2 to 40. Total sample size across θ values is 408,478 samples of size 30.

Using Mathematica 8.0 with MathStatica 2.5 installed, I calculated expected values for each θ level used in simulation, employing Equation (2.5). Table 2.1 compares the expected and observed values. In all cases, the analytical results are extremely close to the observed mean K_n values from simulation, and certainly well within 1 standard deviation. Thus, I conclude that the **TransmissionFramework** implementation of WF-IA employed in this study accurately represents the desired theoretical model.

2.4.2 Time-Averaging and Simulation Parameter Space

Time-averaging was modeled in **TransmissionFramework** by implementing a series of statistical "windows" within which trait counts were accumulated between time steps. At the end of each temporal window, a sample of appropriate size is taken from the accumulation of trait occurrences, trait counts within that sample tabulated, and K_n values recorded. The simulator architecture allows an arbitrary number of temporal windows to be employed simultaneously (albeit with a small performance penalty for each window). As a consequence, during a single simulation run, the simulator tracks both unaveraged statistics and the same statistics averaged over any number of "assemblage durations." All trait samples taken in the simulator, whether unaveraged or for a specific assemblage duration, were also recorded to allow calculation of Slatkin's Exact test. Additionally, to facilitate analysis of time scales within the simulation model, for each trait the interval between entry and loss through drift was recorded. In the simulation results reported here, trait samples were of uniform size 100. Constant sample size removes the effect of different sample sizes on the reported results, although the interaction of the fixed sample size and the innovation rate will lead to cutoff behavior at very high θ values. This is acceptable since the very highest θ values employed here are unrealistic for almost any prehistoric phenomena, and may be approached only for "viral" behavior on modern social networks.

All simulations reported here were performed with a population size (N) of 2000 individuals, and simulation runs were conducted for the following values of θ: 0.1, 0.25, 0.5, 1.0, 2.0, 5.0, and 10-100 at intervals of 10. This range encompasses innovation rates that are very small, through populations in which a full 5% of the population has a never-before-seen variant each generation. Simulations were performed in several batches, with a core set of runs performed for 40,000 steps in order to determine the effects of long-duration time averaging, yielding simulated assemblages at a variety of windows ranging from 3 steps to 8000 steps (the exact durations sampled are given in the first column of Table 2.2). In order to increase the sample size of long-duration assemblages, a second set of simulation runs using the same parameters were done with only the five largest windows recorded (the short duration window sizes were discarded to avoid a flood of raw data beyond that needed for stable statistical analysis). Finally, since the statistical behavior of the process at very small values of θ is highly variable, a third set of runs was performed to increase the number of samples for θ values between 0.1 and 1.0.

Trait samples were post-processed outside the simulator environment, since calculation of Slatkin Exact tests within the simulator itself would slow down the primary simulation model by a large factor. Montgomery Slatkin's original C language program was used in Monte Carlo mode to produce an estimate of $\mathbb{P}(E)$ for each sample of individuals. I modified Slatkin's original `montecarlo.c` program to not require the data to be embedded in the source code, instead taking data as a command line parameter, and outputting only the $\mathbb{P}(E)$ value and θ estimate, to allow easy post-processing of the simulation output.[10]

The simulation results reported here, once post-processed, comprise 3,024,085 sample values for K_n, across the θ values listed above, and broken down across assemblage durations as in Table 2.2, and 1,113,134 Slatkin Exact test results for the same combinations of θ and assemblage duration.

2.5 Results

Simple inspection of the relationship between assemblage "duration" (i.e., accumulation interval) and the average number of variants (K_n) in a sample of size 100, shows a strong time averaging effect (Figure 2.1).[11] Temporal aggregation of the results of transmission inflates the number of variants we see in a sample, with greater effect as the population innovation rate (θ) increases. The effect is very small at low theta values (i.e., when the process is drift-dominated, $\theta < 1.0$) and requires long accumulation of copying events to have a measurable effect upon mean K_n. Conversely, inflated K_n

---

[10] These modifications are available, along with all other analysis scripts, in the Github repositories http://github.com/mmadsen/saa2012, and the **TransmissionFramework** source code.

[11] Here, the time axis represents raw simulation steps, each of which represents $N = 2000$ copying events within the population. This is the only figure in this paper which uses raw simulation time steps as the time variable.

TA Duration	Min Sample Size	Max Sample Size
1	130494	247491
3	4497	43494
7	1926	18639
15	897	8694
31	435	4209
62	216	2103
125	105	1038
250	516	981
500	255	486
1000	114	228
2000	57	114
4000	27	54
8000	12	16

Table 2.2: Breakdown of sample sizes for analysis of trait richness (K_n), by size of time-averaging "window." Some values of θ required larger numbers of simulation runs to achieve stable result, thus the difference between samples sizes at the same TA duration.

appears at fairly short duration as theta increases.

Simulation steps (or "generations") represent an arbitrary time scale with respect to the chronological time archaeologists can (with effort) measure. In order to understand the effects of time averaging on archaeologically-relevant time scales, it will be useful to rescale simulation time by some factor which is observable as a function of artifact class duration in the depositional record. I take up this issue further in Section 2.6, but the ideal time scale would be the mean duration of artifact classes in the classification system being used in a given empirical study. I do not explicitly model archaeological classification in the present results, but a related measure is the lifespan of the traits being transmitted within the simulated population.

2.5.1 Time Scales and Time averaging

The "mean trait lifetime" in WF-IA is a direct consequence of the balance between innovation and loss of traits to drift, in a fixed-size population. At the quasi-stationary state, the population will fluctuate around a mean number of traits, as individual traits enter and leave the population constantly. This implies that at stationarity, if we add traits at a higher rate due to migration or innovation, more traits must be lost to drift each generation. WF-IA thus satisfies a balance equation characterizing the average number of variants ($\bar{n}$)(Ewens, 1964):

$$\frac{\bar{n}}{\bar{t}} = \theta \tag{2.9}$$

where $\bar{t}$ represents the average number of generations that a new trait lasts in the population before its loss to drift (i.e., the mean trait lifetime).

An exact expression for mean trait lifetime has not been derived from the transition probabilities of the WF-IA Markov chain (Ewens, 1964), but it can be approximated by summing the average amount of time that a trait within a population spends at specific frequencies (i.e., mean sojourn times). Ewens (2004, Eq. 3.20) gives the following approximation:

$$\bar{t} \approx \mathbb{E}(t_i) = \sum_{j=1}^{\infty} \frac{2N}{j(j-1+\theta)} (1-(1-p)^j) \tag{2.10}$$

Since θ is in the denominator of the summation, increasing the population rate of innovation reduces the mean trait lifetime by decreasing the amount of time any specific trait spends at a given frequency, and thus the total amount of time a trait spends in the population before being lost to drift.

Table 2.3 lists the observed mean lifetime of traits for each level of θ employed in this study, and the expected value as calculated using Equation 2.10. The observed values are systematically lower than the expected values, which reflects slightly faster loss of traits due to drift in a finite and

Theta	Mean Trait Lifetime	$\mathbb{E}(t_i)$
0.10	36.54	36.89
0.25	25.61	24.05
0.50	21.10	19.97
1.00	17.31	17.21
2.00	14.57	15.21
5.00	12.43	13.05
10.00	10.83	11.57
20.00	9.50	10.16
30.00	8.68	9.36
40.00	8.12	8.79
50.00	7.72	8.36
60.00	7.36	8.01
70.00	7.08	7.72
80.00	6.83	7.46
90.00	6.60	7.42
100.00	6.40	7.05

Table 2.3: Mean lifetime (in model generations) of traits, by θ, along with analytical approximation from Equation 2.10.

small population compared to the large populations often studied in population genetics (Ewens, 1964; Kimura and Crow, 1964). Examination of Figure 2.1 appears to show that the onset of time averaging effects, however small, occurs around the time scale of the mean trait lifetime, for values of $\theta \geq 1.0$. This outcome is sensible given the enhanced probability of longer duration samples incorporating new variants in the sample due to innovation. In the analyses to follow, I scale the time variable by the mean trait lifetime, displaying assemblage duration as a multiple of this value. Thus, for the remainder of this paper, a scaled assemblage duration of 100 will indicate 100 times the mean trait lifetime at that specific θ value. For example, if we are examining results at $\theta = 5.0$, a scaled duration of 100 would indicate $12.43 * 100 = 1243$ simulation steps.

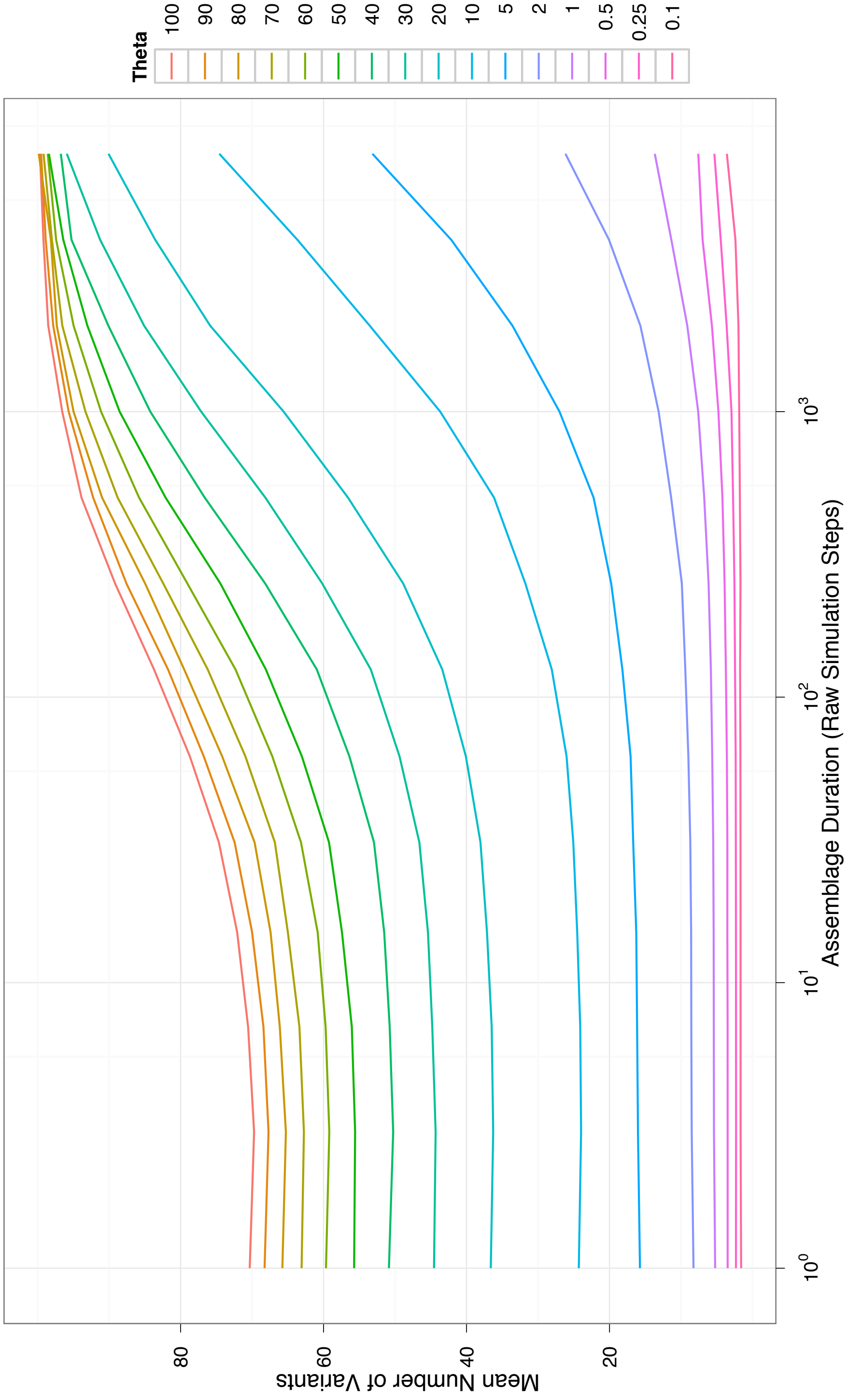


Figure 2.1: Mean value of K_n for time averaged samples, plotted against assemblage duration in simulation steps, for each level of θ in the study. Note that the "onset" of time averaging effects (as measured by increased K_n), is quite gradual at low θ, while high innovation rates display increased richness with very minor amounts of time averaging.

2.5.2 Neutrality Testing

The Slatkin Exact test for neutrality, discussed in Section 2.3.1, determines the "tail" probability for a sample of size n, with observed number of traits k, to be derived from the Ewens Sampling Formula (Equation 2.2). The test employed in this study is Slatkin's Monte Carlo version, which allows the use of larger sample sizes, using random selection to create unlabeled configurations from the ESD to compare against the observed values. The resulting tail probability is converted into a standard hypothesis test by selecting an α value. For purposes of this study, I considered the upper and lower 5% of tail probabilities to indicate that a sample was probably not derived from a neutral transmission model, leading to $\alpha = 0.10$.

Given this α level, we should expect roughly 5% of the samples taken from a pure neutral copying process to fall into each of the the upper and lower tail regions, and thus for a Slatkin Exact test to reject the null hypothesis of neutrality. Roughly 90% of the samples we take from the neutral WF-IA process should fall between $0.05 < p < 0.95$ and thus lead to acceptance of the null hypothesis. This experimental setup also implies the limited utility of performing a single neutrality test on a single sample of class counts or frequencies, as has been archaeological practice by necessity. A single Slatkin exact test with $\mathbb{P}_e$ value of, say, 0.96, would constitute some, but relatively weak, evidence of non-neutrality. Better practice would be taking many samples from a large assemblage or multiple collections and calculating independent Slatkin tests for each sample, and examining their distribution.

If time averaging has no effect on the validity of the Slatkin Exact test employed against temporally aggregated samples, we would expect the fraction of samples in the two tails (upper and lower 5% in this case) to equal 10%. Anything over 10% would constitute evidence of extra Type I errors, since we know the samples to have been generated by a process meeting the definition of the null hypothesis. Therefore, after post-processing the simulation output to produce Slatkin tests as described in Section 2.4, I tabulated the fraction of Slatkin Exact tail probabilities that exceeded the expected 10% tail population. These are, in other words, "excess" failures of the Slatkin Exact test, beyond those expected by the probability distribution itself. For each θ level, and for each time averaging duration, the mean "excess" failure rate was computed, from the 1,113,134 raw Slatkin Exact test results generated in the simulation study.

Figure 2.2 depicts the relationship between the excess failure rate, and time averaging duration scaled by the mean trait lifetime (as previously described). The mean trait lifetime is indicated by a vertical red bar in each graph. Three major results are apparent. First, at values of $\theta \geq 1.0$, the excess failure rate in non-time-averaged data is zero, as one would expect, and then begins to increase (albeit slowly) as the time averaging duration of samples exceeds the mean trait lifetime. In some cases, such as $\theta = 5.0$, the Slatkin Exact test continues to be accurate given the chosen α value through samples which are aggregated for 10 times the mean trait lifetime. But in all cases, with sufficient time averaging, the Slatkin Exact test begins to suffer from increased Type I error, reporting an ever increasing fraction of samples as not derived from a neutral transmission process. The extreme situation is seen at very high rates of innovation, where nearly every test fails, at high levels of time averaging. These failures are caused by saturation of a finite sample with singleton traits, causing the sample to display too much evenness in frequency to have been the result of sampling from the Ewens Sampling Formula. But unrelated to this saturation effect, there is considerable failure in employing the Slatkin Exact test to detect neutrality. For example, at $\theta = 5.0$, at 100 times the mean trait lifetime, approximately 70% of all samples appear in the tail region of the distribution, compared to the expected 10%. Clearly, the Slatkin Exact test is not robust for long-duration assemblages.

Second, at low θ values, the test results show excessive Type I error, even without time averaging. There are several potential causes. It is possible that the WF-IA process had not reached quasi-stationarity by 750 time steps, when sampling began. This would mean that the effects of initial

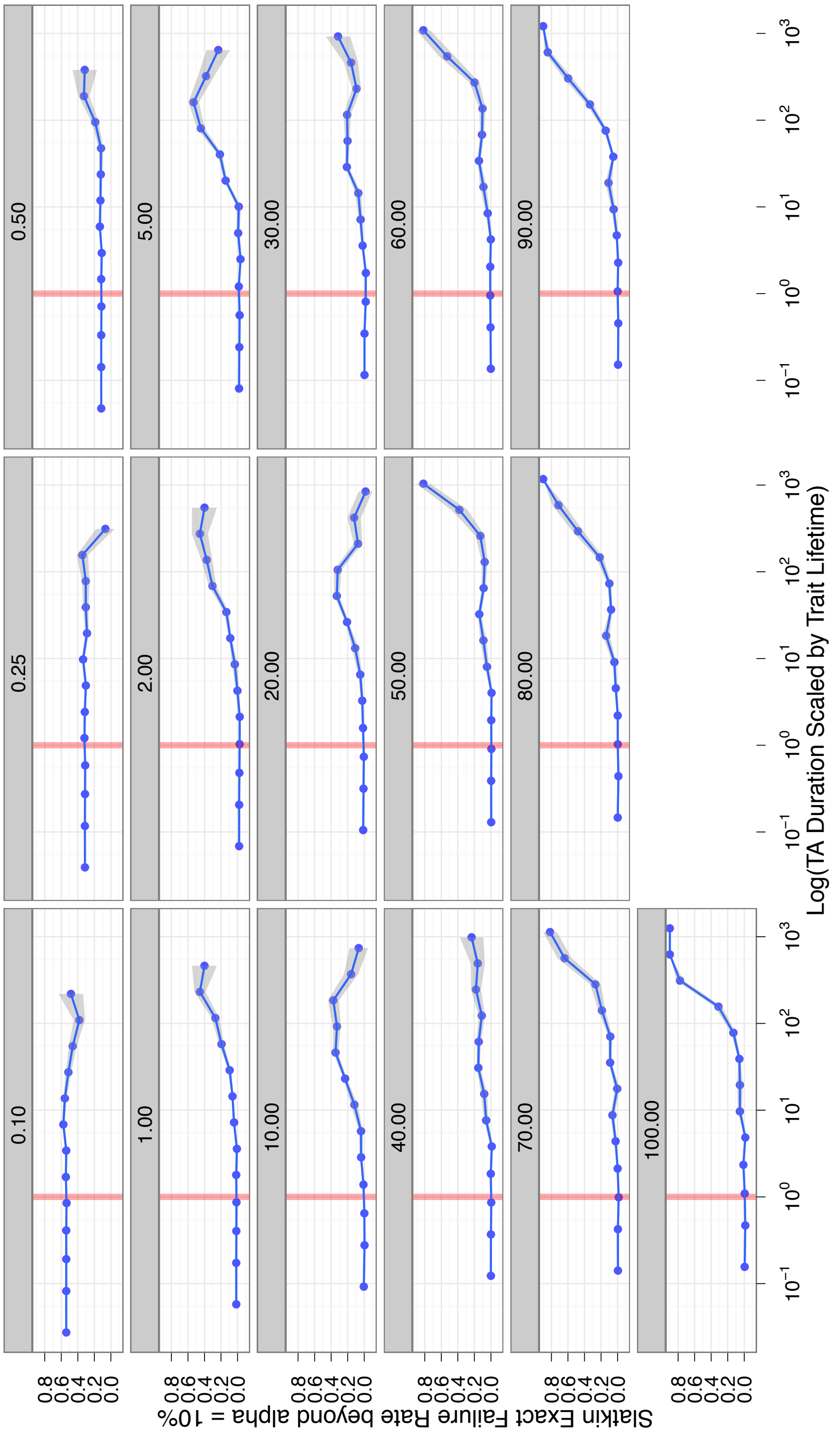


Figure 2.2: Slatkin Exact test failure rate (above the expected 10% given two-tailed test with $\alpha = 0.10$, plotted against time averaging duration scaled by mean trait lifetime, for each level of θ in the simulation study. The red vertical line indicates the mean trait lifetime for that θ value, and the shaded region encompasses the standard error of the estimates for mean failure rates at each duration.

trait assignment might still be present and skewing the frequency distribution of traits. Second, the Slatkin test is sensitive to the number of rare or singleton traits given the sample size, and in a small population (2000 individuals) with a low innovation rate (e.g., $\theta = 0.1$), counts of rare traits could be unstable. This would not typically be the case in samples from large populations or entire species. I do not consider the cause of this anomaly further in this paper, but it warrants further simulation study.

In general, with long-duration assemblages, archaeologists should be careful interpreting the results of neutrality tests adopted from population genetics. The effect seen here can be summarized as: with significant time averaging, trait frequencies generated by unbiased cultural transmission can falsely appear to be non-neutral and thus driven by bias or selection (Type 1 error). The longer the duration of an assemblage with respect to the mean trait lifetime, the larger the probability of a Type 1 error. With sufficient duration, in fact, the probability of a Type 1 error becomes virtually certain, and the Slatkin Exact test loses any discriminatory power. In summary, if one were to employ Slatkin's test to examine the hypothesis of neutrality in long-duration archaeological deposits, one would overwhelmingly come away with the impression that most cultural transmission was biased, either towards conformity or a pro-novelty bias – regardless of the underlying process occurring during prehistory.

2.5.3 Theta Estimation and Innovation Rates

There would be considerable value in estimating the population-level innovation rate (θ) from sample data if it could be done accurately. As discussed in Section 2.3.2 above, such estimates are usually biased and have large variance. In this section, I examine the effects of time averaging upon theta estimates generated from the samples taken to perform neutrality tests in the previous section. For each of the 1.1 million samples of variants (distributed across actual theta values and assemblage durations), I calculated theta estimates given Watterson's approximation (Durrett, 2008):

$$\hat{\theta} \approx \frac{k_n}{\log n} \tag{2.11}$$

For each combination of actual theta and assemblage duration, theta estimates were averaged, to give a mean estimated theta value ($\mathbb{E}(\hat{\theta})$), and its standard deviation. The results are shown in Table 2.4. There are two regions of behavior apparent in the table, corresponding to drift- versus innovation-dominated dynamics. At and below $\theta = 1.0$, estimated values are higher than the actual θ used to generate samples, and above 1.0, theta estimates begin to systematically lag below the actual theta value. Overestimation at $\theta \leq 1.0$ matches the simulation results by Ewens (1974), although the authors did not simulate innovation rates above 2.0 (a large value in most genetic situations). In addition to being biased, theta estimation appears to be even *approximately* accurate only within a narrow range of values around $\theta = 1.0$.

Figure 2.3 examines estimates of theta by time averaging duration scaled by the mean trait lifetime, for each level of actual θ used in the simulation runs. The pattern evident in synchronic or unaveraged samples carries over to time averaged assemblages: below $\theta \leq 1.0$, theta estimates are larger than the actual values, and increase in a non-linear fashion with assemblage duration. Above 1.0 but below about 30.0, theta estimates begin below the actual value, cross the actual value, and continue to accumulate as assemblage duration increases. Finally, at the very highest innovation rates, in a sample size 100, theta estimates are always drastic underestimates of the actual value, even with long assemblage duration increasing the accumulation of traits.

The Slatkin Exact test software also provides an estimate of θ, finding the maximum likelihood value of theta when K_n is set in Equation 2.5 to equal the observed value (this is the t_e statistic introduced to archaeological usage in Neiman, 1995). Figure 2.4 depicts the Slatkin theta estimates by time averaging duration scaled by the mean trait lifetime, for each level of actual θ used in the

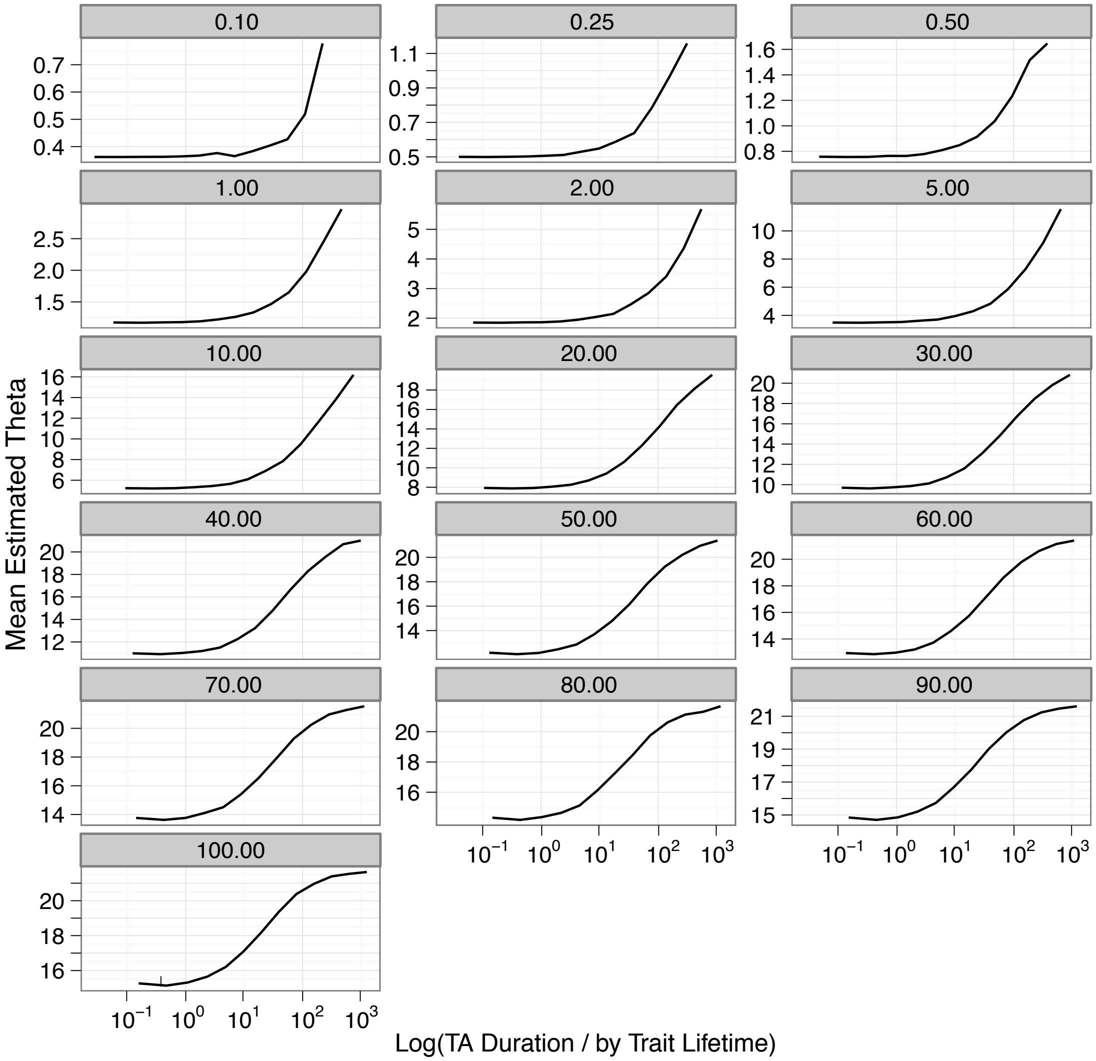


Figure 2.3: Estimates of mean population innovation rate ($\mathbb{E}(\hat{\theta})$) from samples ($n = 100$) taken for neutrality tests, using the approximation by Watterson (1975). Plotted against assemblage duration, for each level of actual innovation rate used in simulation runs.

θ_0	$\mathbb{E}(\hat{\theta})$	$\sigma(\hat{\theta})$
0.10	0.36	0.21
0.25	0.50	0.26
0.50	0.76	0.33
1.00	1.17	0.42
2.00	1.85	0.51
5.00	3.49	0.67
10.00	5.23	0.87
20.00	7.93	0.95
30.00	9.70	0.99
40.00	10.99	0.99
50.00	12.19	1.00
60.00	12.94	1.01
70.00	13.76	0.97
80.00	14.32	0.98
90.00	14.85	0.94
100.00	15.27	0.95

Table 2.4: Mean Estimated Theta ($\mathbb{E}(\hat{\theta})$) from Samples (n=100) compared to actual values employed in simulation models (θ_0), without any time-averaging.

simulation runs. One interesting difference between Figure 2.3 and the Slatkin theta estimates is that the latter are more accurate for actual $\theta \geq 1.0$ than the Watterson approximation, in unaveraged assemblages. Unfortunately, with increased assemblage duration, estimates explode to much larger values than those calculated by the Watterson approximation (i.e., $\theta \approx 1500$ for true $\theta = 30$ at maximum assemblage duration of 1000 times the mean trait lifetime, compared to the *underestimate* of approximately 22 in Figure 2.3).

In short, estimation of population-level innovation rates from samples of artifacts using either estimation method are inaccurate, and the time averaging effect of accretional deposition renders such estimates even more inaccurate. Clearly, such values cannot be used as actual indications of innovation rate or to "work backward" towards past population sizes. And without fairly precise control over assemblage duration, the use of t_e as a relative diversity measure between assemblages

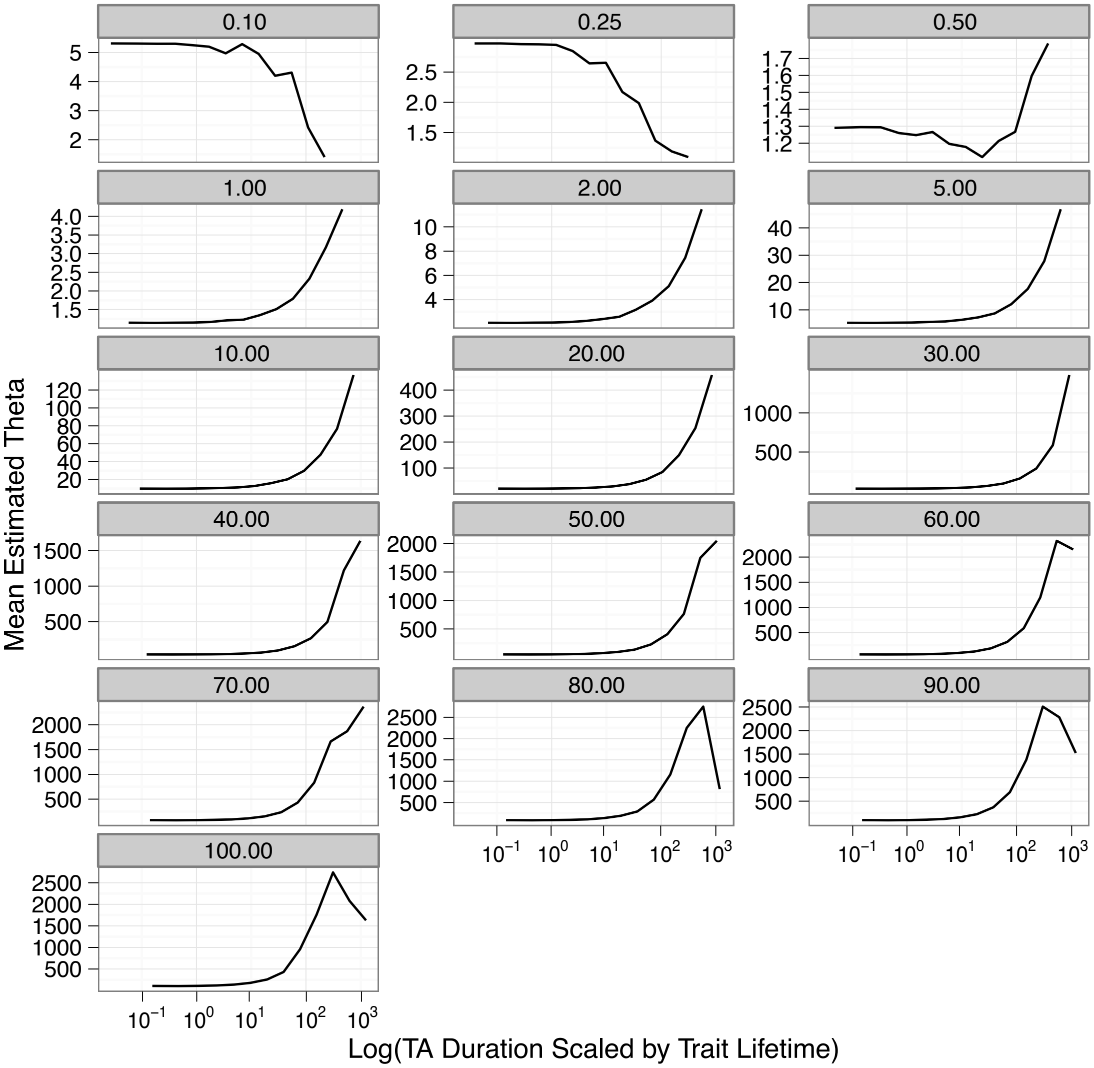


Figure 2.4: Estimates of mean population innovation rate ($\mathbb{E}(\hat{\theta})$) from samples ($n = 100$) taken for neutrality tests, using results from Montgomery Slatkin's neutrality test software. Plotted against assemblage duration, for each level of actual innovation rate used in simulation runs.

(in the manner common to archaeological applications) is highly suspect. In the next section, I turn to t_f, the other common diversity measure in archaeological studies, which does not require an estimate of θ, employing instead the variant frequencies directly.

2.5.4 Diversity Measures

Much of the current effort in distinguishing biased and unbiased transmission models rely upon trait evenness and the shape of frequency distributions, given Alex Bentley's application of power-law distributions to both ancient and contemporary data sets (Bentley and Shennan, 2003; Bentley et al., 2004; Bentley, 2007; Bentley et al., 2009; Hahn and Bentley, 2003; Herzog et al., 2004). One of the ways that unbiased and "conformist" models of cultural transmission differ is in the expected amount of variation. Compared to unbiased transmission, conformism of even a mild degree tends to strongly concentrate adoption onto a very small number of traits (Mesoudi and Lycett, 2009).[12] It is difficult, however, to interpret the absolute number of traits (K_n) without knowledge of the population size, so Kohler et al. (2004) employed diversity measures instead in his classic examination of conformist transmission in Southwest pottery.

The most commonly used measure in the archaeological literature on cultural transmission is t_f (Equation 2.6), since it is related to Wright's original measures of heterozygosity and thus associated directly with the historical development of the Wright-Fisher model. But it is useful to normalize the results of t_f between 0 and 1 so that we can compare different levels of theta and assemblage durations easily, in the same way that statisticians occasionally employ coefficients of variation or normalize covariances into correlation coefficients. Equation 2.7 does exactly this, and is called the "index of qualitative variation" (IQV) (Wilcox, 1973).

---

[12] This is especially the case when conformist transmission is implemented in simulations as a "global" rule where only the most common trait is copied during "conformist" copying events, rather than weighting all traits by their relative popularity. Very little work has been done to compare the results from different methods of simulating biased transmission models. This is a topic which would benefit greatly from additional research.

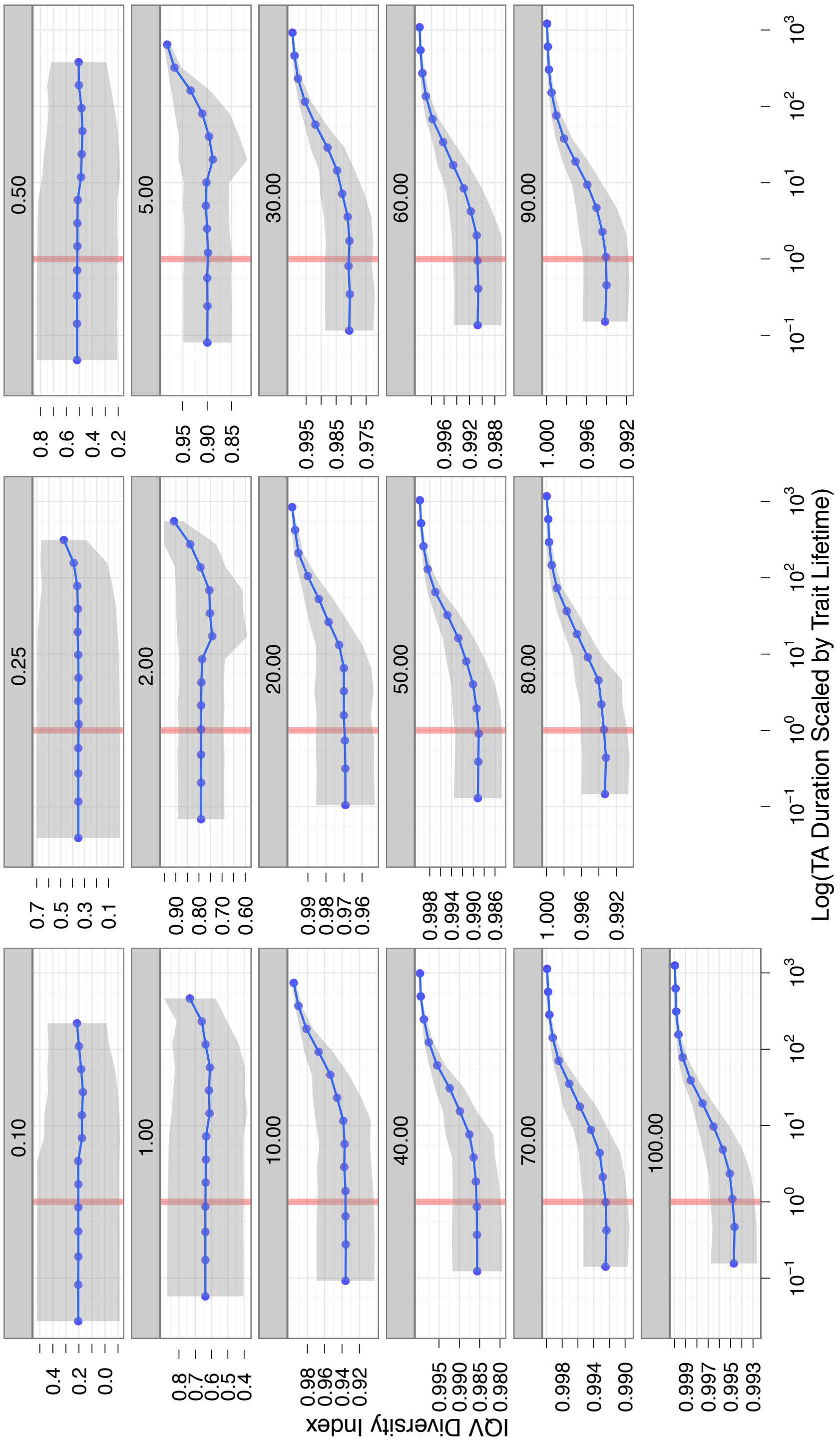


Figure 2.5: IQV diversity index, derived from samples of size 100, plotted against time averaging duration scaled by mean trait lifetime, for each level of θ in the simulation study. The red vertical line indicates the mean trait lifetime for that θ value.

Figure 2.5 displays the relationship between the IQV for samples of size 100, and time averaging duration scaled by mean trait lifetime, as before. IQV values range from 0.0, if only a single trait occurred within a sample (which happens in simulations with very low innovation rates), through 1.0, which indicates that traits are perfectly evenly distributed within a sample. Even at the highest innovation rate studied, values of 1.0 were not seen in *unaveraged* samples from the simulation runs. It is apparent that time averaging can yield greater evenness among trait frequencies, although the plateau in IQV values seen at high θ and high assemblage duration is a function of the saturation of K_n in a finite sample seen above. At very low innovation rates ($\theta \ll 1.0$), time averaging in contrast seems to have little effect on the dispersion of trait frequencies, with one or a very few traits always dominating a sample.

In between, when innovation rates are sufficient to guarantee at least one innovation on average per model generation ($\theta = 1.0$) but fewer than 10, there is non-monotonic behavior apparent in the IQV index. For example, at $\theta = 2.0$, time averaging has no effect on IQV until duration is 10 times the mean trait lifetime ($\bar{t}$), at which point assemblages begin to appear *less even* in frequency distribution, until about 100 times the mean trait lifetime, when evenness begins to steadily increase. This effect is interesting, since it suggests that we cannot easily compare diversity indices between assemblages unless we control for duration or have independent evidence concerning innovation rates.

2.6 Discussion and Conclusions

When we examine the effects of time averaging on the sample properties of unbiased transmission, using the mean lifetime of traits as our fundamental time scale, several lessons for practical applications emerge. First, it appears that assemblages with very small amounts of temporal aggregation display little of the distributional alterations that characterize long-duration assemblages. Statistical tests of neutrality and diversity measures, and thus arguments based on them, can probably be used with care. Second, estimates of population-level innovation rate derived from Ewens's sampling formula are biased (and therefore inaccurate), and become seriously inaccurate with increased assemblage duration. Archaeologists should strongly reconsider using t_e or other theta estimates even in relative comparisons, and should definitely not consider such estimates to reflect the innovation rate or population size present in the prehistoric population. Third, for assemblages that have a duration longer than the mean trait lifetime, it is important to measure and control for the relative duration of assemblages when comparing statistical results across samples. Without doing so, we cannot interpret relative differences of diversity indices or trait richness values as indicative of different modes of transmission.

One caveat to the above is that such effects refer specifically to *assemblage* level data, composed of many artifacts deposited over time. Artifact-scale analysis, where the attributes under analysis come together in a short period of time, and where single artifacts comprise the counting unit for transmission studies, will not necessarily suffer the quantitative effects described here, or would suffer no measurable time averaging effects if the assemblage durations were short compared to the lifetime of traits. A good example of this is Jonathan Scholnick's chapter in the present issue, expanding on his previous research into cultural transmission in Colonial America through gravestones (Premo and Scholnick, 2011; Scholnick, 2010), where his samples cover 10 year periods based on the death dates carved on each stone.

Furthermore, while the mean lifetime of transmitted information plays a central role in establishing a "natural" time scale over which time averaging affects unbiased transmission, this time scale is not an archaeological one. This discrepancy in time scales arises because the abstract "traits" of our models are not equivalent to the classification units employed by archaeologists. This is not a trivial difference, and is one that is rarely even discussed in archaeological applications of cultural transmission models. Instead, we frequently act as if "traits equal types," despite occasional acknowledgement of the difference.

But we have no direct empirical access to the information prehistoric populations were learning, teaching, and imitating. We will never find "units of transmission" in any empirical sense for archaeological applications of cultural transmission models, and we have no warrant to equate our models of prehistoric information flow with the classes we use to observe it today. Long ago, Osgood (1951) recognized that when anthropologists study the ideas held within a social group under study, what is actually being studied are the ideas we *construct* about the ideas individuals in other cultures may have had. Dunnell (1971) systematized this distinction, pointing out that we always operate with analytic classes whose construction is done by archaeologists, for archaeological purposes. These classes serve as a "filter" by which we detect patterns in artifact assemblages, which reflect patterns in the information which flowed within past populations. There is no "natural" set of classes to employ in studying cultural transmission, but we often forget to incorporate this fact into our analyses. Linking the time scale over which variation entered and left a prehistoric population, and the time scale over which archaeological classes appear and then exit the archaeological record will involve further research on the relationship between transmission dynamics and complex, multi-dimensional archaeological classes. Such research is essential to connect the abstract quantities described by theoretical models, to observable aspects of the archaeological record.

These results paint a fairly gloomy picture of almost all of the standard variables archaeologists have used since Neiman's (1995) pioneering work. One wonders why empirical studies using diversity measures, innovation rate estimates, or neutrality tests appear to "work" and give sensible results? One possibility, of course, is that some studies don't yield the expected results. We see this, possibly, in a fascinating analysis by Steele et al. (2010). The authors employed ceramic classes that appeared to be non-neutral and subject to selection or biased transmission. Yet Slatkin exact tests were unable to rule out the null hypothesis of neutrality. I do not present an analysis of conformist transmission under time averaging in this article, but using **TransmissionFramework** I see evidence that temporal aggregation has the opposite effect on Slatkin exact tests in populations with weak conformist biases: neutrality tests suffer increased Type II error, making it more likely that we will accept a null hypothesis of neutrality when the opposite is the case.

Another possibility is that certain variables may retain their distributional character, but have their values inflated by temporal aggregation. In such situations, there would be no reason to reject the neutral model, but inferences about the values of parameters would be inaccurate. Even if investigators did not rely upon the absolute value of parameters, frequently such inferences (e.g., diversity values) are employed as relative comparisons between assemblages. I suspect that this has occurred in a number of published studies, but few cultural transmission applications include detailed information concerning assemblage duration, so it is difficult to redo the researcher's original hypothesis tests with temporal controls, without going back to original field information or reports. Clearly, both possibilities may also occur in some situations.

As archaeological usage of cultural transmission theory becomes more frequent and we move from proof-of-concept studies to demanding interpretive accuracy from our models, methodological research is essential to ensure that our applications are empirically and dynamically sufficient. The present study focused on a necessary first step in such method development, developing an understanding of the effect of time averaging in accretional assemblages upon the observable variables in neutral cultural transmission models. The results demonstrate that frequently employed statistics, such as t_e, are highly inaccurate and biased when measured in time averaged assemblages, and that neutrality tests are subject to enough additional Type I or Type II error that the results can be systematically misleading. Clearly, in order to apply cultural transmission models to diachronic data derived from time averaged assemblages, we need to develop observational tools and methods suited specifically to the archaeological record, instead of simply borrowing statistical methods and models from theoretical population biology.

2.7 Acknowledgements

This paper was originally presented in a symposium titled *"Recent Developments in Cultural Transmission Theory and its Applications"* at the 2012 Annual Meeting of the Society for American Archaeology, Memphis, TN. The author wishes to thank Kristen Safi, the organizer, for the opportunity to participate, and Carl P. Lipo, Fraser Neiman, James Feathers, Jonathan Scholnick, and Michael J. O'Brien for comments on drafts of this paper.

CHAPTER 3

Can We Identify Biased Cultural Transmission in the Archaeological Record?

3.1 Introduction

The emerging field of cultural evolution is the study of cultural change in humans and other animals as a Darwinian evolutionary process, and encompasses research in biology and the social sciences. As part of an "extended synthesis" (Pigliucci and Müller, 2010), cultural evolution extends the notion of "descent with modification" to include the learning and accumulation of cultural information between generations, and explicit study of the differences in structure and pattern between genetic and cultural transmission. Beginning with the seminal work by Cavalli-Sforza and Feldman (1981) and later by Boyd and Richerson (1985), the study of these differences has focused a great deal of attention on cognitive and psychological biases that create population structure. Over the past thirty years experiments, observational studies, and theoretical models have come together to create a picture of human social learning which is biased towards conformity under many circumstances and the observation of success, positive payoffs, and social prestige as a basis for selecting those from whom one learns or imitates (e.g., Richerson and Boyd, 2005).

Most of the evidence for this emerging picture of human social learning comes from controlled experiments and observational studies of living populations. Most of it is conducted by psychologists and anthropologists (Whiten et al., 2016; Mesoudi et al., 2016). Archaeologists and paleoanthropologists seeking to study the evolutionary history of social learning are consumers of this body of theory. There is nothing wrong with being a consumer of evolutionary theory—every discipline involved in applying Darwinian insights to human social behavior is a consumer in some area—but we must also be producers of methods needed apply that theory to the unique empirical phenomena we study. Our job consists, at a high level, of building methods for constructing models appropriate to the kinds of data we possess, understanding how to best to assess the fit between those models and our data, and finally, understanding the limits of our ability to do so given the nature of the empirical phenomena we study.

This paper addresses the limits of our ability to select among detailed, individual-level models of cultural transmission in most archaeological situations. Initial optimism applying mathematical models of cultural transmission to archaeological data (e.g., Neiman, 1995; Eerkens and Lipo, 2005; Lipo et al., 1997; Shennan and Wilkinson, 2001; Jordan and Shennan, 2003) has given way to a more nuanced view of our ability to discriminate between models (Barrett, 2019; Premo, 2010; Kandler and Crema, 2019). A growing body of work is aimed at assessing the causes of equifinality between cultural transmission models given archaeological data, for example assessing the effects of time averaging (Madsen, 2012; Porčić, 2014; Premo, 2014; Perreault, 2018) and non-stationary population sizes (Rorabaugh, 2014; Kandler and Powell, 2018). The hope is that methodological development will help us understand, and correct for, these sources of equifinality so that we can proceed with "model selection" and fitting (albeit in a more sophisticated way) and still work with the kinds of transmission models that other anthropologists and social psychologists employ (especially those derived from Boyd and Richerson's seminal work on "dual inheritance theory").

There is no question that we can empirically detect different types of social learning modalities in controlled experiments (Kempe and Mesoudi, 2014; Mesoudi et al., 2006a; Whiten et al., 2016; Mesoudi and O'Brien, 2008a; Mesoudi et al., 2008; Mesoudi, 2014). Experiments and detailed observational studies of living populations generate fine-grained data about who individuals learn from and what covariates affect "chains" of transmission. Archaeologists, even in the best of circumstances, always face data which is coarse-grained compared to that available in experimental settings. To one degree or another, the data we generate from archaeological deposits are estimates of some population-level prevalence of cultural traits.

Understanding the limitations of using coarse grained data to perform model selection is crucial for archaeology and paleoanthropology, The main approach to identifying the best fit cultural transmission model for a data set has been to construct univariate statistical tests or distributional expectations for various summary statistics. Examples of summary statistics employed in this model fitting approach include richness or diversity (evenness) measures, the average amount of time traits survive in a population, their turnover properties, or the degree to which frequency data match distributions known to arise in various "null" models (Shennan and Wilkinson, 2001; Shennan and Bentley, 2008; Bentley and Shennan, 2003; Shennan and Bentley, 2008; Kandler and Shennan, 2013, 2015). These statistics are highly attractive for archaeological purposes because most are easy to calculate from standard descriptions of archaeological assemblages without additional measurements on individual artifacts and thus allow the study of previously described data sets.

This approach has now been employed in enough studies, and with enough replications on the now-paradigmatic ceramic data from the European Neolithic Merzbach Valley, that we can evaluate its performance. Kandler (2015) notes that after five separate analyses of the Merzbach data using a variety of summary statistics and approaches, the results remain in conflict. It is not clear whether the data are consistent with a hypothesis of neutrality, or anti-conformity/novelty seeking, or whether both are reflected in different assemblages.

There has been much less focus on the crucial question of whether some of the equifinality we face is simply structural. By "structural" equifinality, I mean that the models we expect to be able to identify cannot be distinguished because they overlap too strongly in their outcomes, with any possible data we can obtain. This paper is an attempt to outline a major cause of structural equifinality in cultural transmission modeling: the inherent variability of social learning modes within real populations. The simulation experiments reported here seek to understand whether individual variability in cultural transmission and social learning strategies renders our models impossible to distinguish given coarse grained data, and what effects that sampling and time averaging have on our ability to distinguish the true data generating process behind our data.

This study employs simulation from several data generating models, and employ multiple summary statistics and a powerful machine learning classifier capable of finding highly nonlinear decision boundaries to determine whether it is even possible to discriminate between our theoretical models. If we can, then it is reasonable to expect that further methodological research on ways to "correct" or avoid sources of equifinality may be fruitful. If we cannot distinguish between models, even in the idealized case, then I believe we must question the utility of attempting to perform the kind of microevolutionary model fitting that many of us have been attempting with archaeological data.

The results indicate that populations with mixtures of bias can be distinguished from a reference population of unbiased copiers very accurately given a full population census and the absence of time averaging. In this sense, equifinality between these theoretical models may not be structural even with coarse-grained data. However, the introduction of sparse sampling and the interaction of sampling with time averaging markedly degrades our ability to accurately classify samples as to their model of origin. Furthermore, the pattern of errors is not symmetric. With sampled, time averaged data, we are extremely likely to conclude that samples represent various kinds of transmission bias, even when this is not the case.

This result deepens the skepticism we should feel as archaeologists about the ability to apply detailed social psychological models to most archaeological data. Instead of attempting to adapt our data so that we can apply microevolutionary models, we may be better off adapting our models to the data and empirical record we actually possess. This means looking at coarser grained models and coarser grained research questions which may better match the unique strengths of a diachronic, aggregated record of human evolution.

3.2 Within Population Variation in Social Learning: A Cause of Structural Equifinality?

The microevolutionary models of cultural transmission we typically employ in archaeology draw structurally upon the mathematical core of classical population genetics, but add inheritance structures and social learning modes informed by observations across the social sciences (Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985). The core models in dual inheritance theory richly vary in the cognitive biases they describe, but all still tend to ignore significant sources of population structure or variation. All tend to assume *panmixia* as a simplifying assumption, and most tend to depict the population-level consequences of sets of individuals who employ a single learning mode or class of cultural "parent" to learn from (Wimsatt, 2019).

Real populations, as opposed to those depicted in much of our modeling, have rich structure. Human populations are mixtures of people of different ages, genders, and propensities for individual versus social learning. There also appears to be variation among populations and major cultural lineages in these mixtures of social learning traits. Detailed experimental and observational studies are just beginning to document this variation (Whiten et al., 2016; Yaveroglu and Donthu, 2002; Tweed and Lehman, 2002; Chang et al., 2011; Enquist et al., 2007; Csibra and Gergely, 2009; Tomasello, 2016; López et al., 2010; Berl and Hewlett, 2015; Mesoudi et al., 2015, 2016) People can vary in these propensities over the course of their lifetimes (Lehmann et al., 2013; Demps et al., 2012; Correa-Chávez and Rogoff, 2009; Mesoudi et al., 2016), making demographic factors important in modeling the population-level effects of individual social learning processes.

Even if different ways of learning from parents and peers are distinguishable from individual level data, we should not automatically assume that their population level consequences are distinguishable, when there is variation in the population. Processes which are distinguishable with transmission chains, for example, may have opposing effects on population-level summary statistics. A population comprised of conformists who prefer to learn and stick with the ways that have worked in the past would tend to display fewer variants in the population than expected in a neutral equilibrium, and more concentration of frequency on a small number of traits. A population comprised of those who prize novelty and exploration of "new" traits will tend to display more variants than expected, with more evenness among trait frequencies. A real population that combines individuals with both propensities would display a mixture of these population-level consequences. In other words, some of the effects easily visible at the individual level may "cancel out". If that is the case, can we distinguish a mixture of these social learning biases, from a population of unbiased copiers?

This question becomes even more concerning when we combine the "structural" possibility of equifinality just described with the "methodological" sources of equifinality previously discussed: sampling and the time averaging characteristic of aggregated, time transgressive data. This question

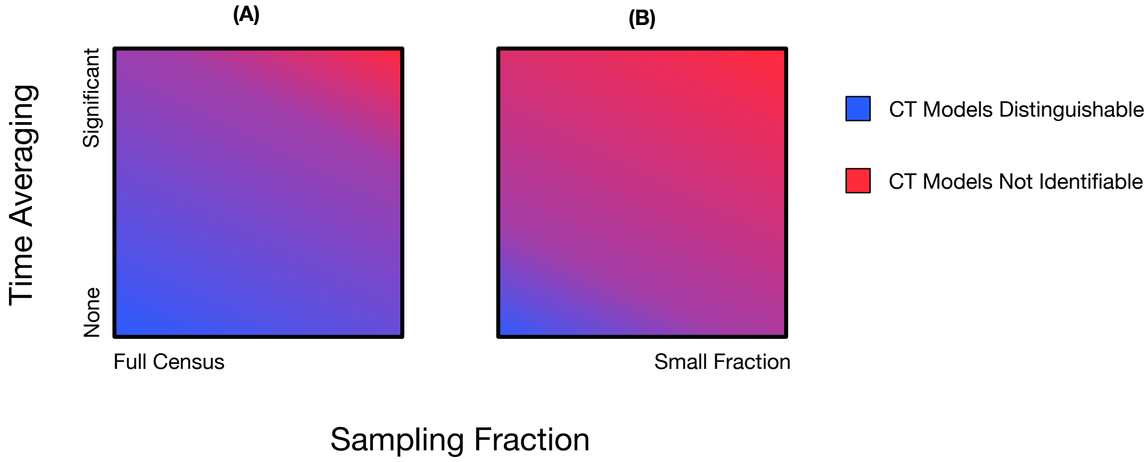


Figure 3.1: Two scenarios for how equifinality may affect our ability to empirically distinguish between different models of cultural transmission given the effects of sampling and time averaging.

is depicted schematically in Figure 3.1. In the left panel, under mild to moderate amounts of time averaging and with reasonable sample fractions, our ability to discriminate between models may be relatively strong, with equifinality restricted to situations with small sample size and in assemblages with significant duration. The right hand panel depicts the other end of a continuum of possibilities: our ability to identify transmission models given data might be quite rare across a range of values relevant to archaeological inquiry. This paper is an attempt to determine where we might stand between these two poles. Given even simplified models of mixed social learning processes, can we tell apart their population level consequences, especially with limited samples and in the presence of time averaging?

3.3 Methods

3.3.1 Study Design

I approach this question by simulating populations with mixtures of social learning or transmission biases. In particular, I examine populations comprised of mixtures of conformist and anticonformist individuals, along with a "reference" population of unbiased copiers (Boyd and Richerson, 1985). The simulated populations are censused at sampling intervals and trait counts recorded. This allows us to later perform a variety of "data collection regimes" which simulate the effects of archaeological sampling and the time averaging effects of aggregate deposition in the archaeological record. It also allows analysis at the single trait level, or the composition of traits into "classes" which simulate the effects of archaeological classification, since most efforts to fit cultural transmission models to real data have accepted existing classification schemes as input to the model fitting.

The output of the simulation runs are simulated archaeological observations of trait or class frequency count at a variety of levels of sample size and time averaging "treatments" from the same underlying data generating processes. This allows us to examine our ability to distinguish between data generating processes depending upon the "treatment" levels, given a method for predicting which data generating process produced the data points observed. In the next section, I describe a predictive modeling approach, adopted from machine learning, for measuring equifinality as error in properly predicting the known data generating process for our simulated data.

3.3.2 Measuring Equifinality Through Classification Error

The common models of cultural transmission employed by archaeologists are stochastic in nature, and thus when we take samples from simulations of those models, we will observe a distribution of results for any summary statistic we choose to observe. To the degree that several models generate separable distributions for the chosen summary statistics, we will be able to infer the most probable data generating model given values of those statistics. If the distributions of summary statistics have some overlap, the level of certainty with which we can predict the correct data generating model will decline overall, and perhaps be no better than chance in the region of overlap.

This generic situation is shown schematically in Fig. 3.2. Here, three pairs of probability models are represented by 500 measurements each of two continuous predictors variables (e.g., a diversity index). In the left panel, the pair of models do not overlap in their outcomes. Given a data point, we can assign it to Model 1 or Model 2 with virtually no error, and thus we would consider models 1 and 2 to be distinct and not equifinal at all. The situation in the middle and right panels of Figure 3.2 is different. There is some overlap in the middle panel, and very strong overlap in the right panel. In the right hand panel, in fact, there is enough overlap that on average, our ability to assign a randomly chosen data point to the correct model is no better than chance. Intuitively, we would say that there is some equifinality in the middle panel, and that the two models in the right hand panel were strongly equifinal.

Measuring the overlap between the measured outcomes of theoretical models thus provides a way to determine their inherent or structural equifinality given a set of observable summary statistics or observational variables. When theoretical models are simple enough to have solvable equations, it may be possible to perform an analysis of distributional overlap using mathematical tools such as the Kullback-Leibler divergence between their resulting probability distributions (Burnham and Anderson, 2002). In the kind of cases discussed in this paper, with mixtures of social learning modes in a single simulated population, we need a numerical and thus statistical approach. We will operationalize measurement of distributional overlap by producing samples from each theoretical model, calculating or other producing the observable variables or summary statistics, applying any simulated data collection treatments (such as specific sampling regimes), and then attempting to

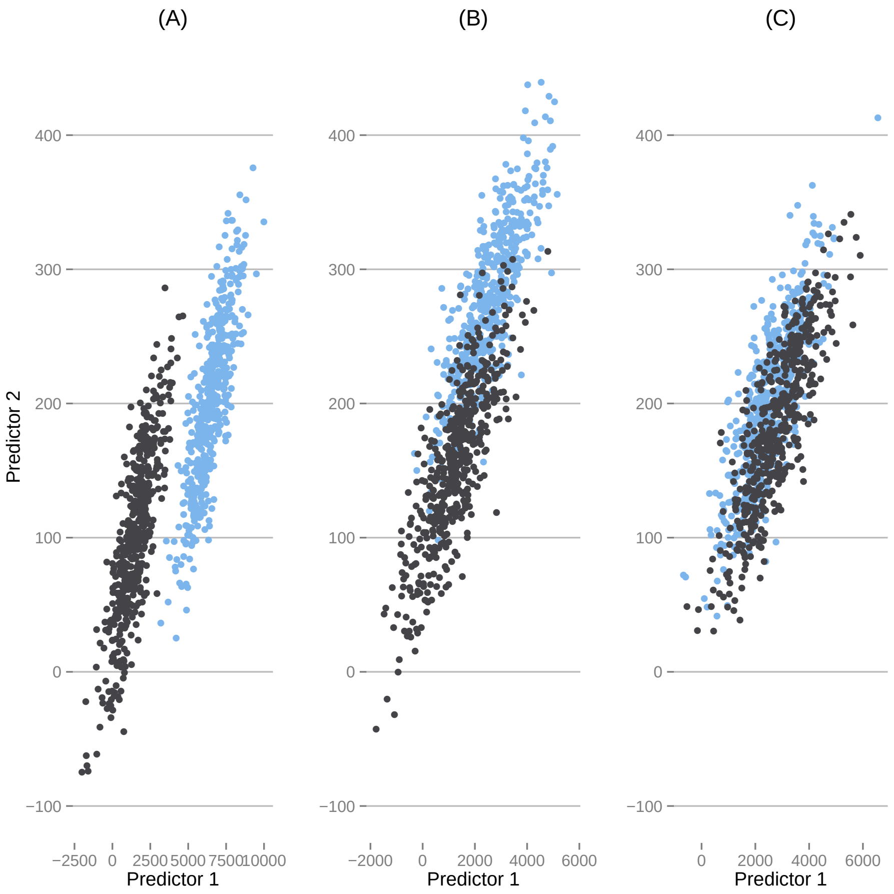


Figure 3.2: Simple example of model outcomes with different degrees of distinguishability given two summary statistics: (A) simulated data points from two fully separate models, (B) two models with a limited overlap region, (C) and two models whose outcomes are highly overlapping.

"predict" which data generating model generated each data point. The error rate in predicting the correct model is a quantitative measure of *how much equifinality* there is between models, given a set of predictor variables and data collection treatments.

In more formal terms, we define measurement of equifinality as the error in performing a classification task, in the machine learning sense (Murphy, 2012). Given a set of models $\mathcal{M}_1 \ldots \mathcal{M}_n$, we can measure equifinality as the minimum possible error achievable in correctly assigning simulated data points to the data generating model which produced them, given measurement of a set of predictor variables or "summary statistics." In the classification task, we ask which model has the highest probability for a given data point, given the conditional density of the data and models. This sounds exactly like Bayes' theorem, and in fact we can write the classification problem as follows, where $Y \in 1, \ldots, K$ refers to each of k models, and $X_1, \ldots, X_p$ refer to p different predictor variables.

$$\mathbb{P}(Y|X_1, \ldots, X_p) = \frac{\mathbb{P}(Y_i)\mathbb{P}(X_1, \ldots, X_p|Y)}{\mathbb{P}(X_1, \ldots, X_p)} \tag{3.1}$$

$\mathbb{P}(Y)$ plays the role of the prior distribution, and reflects how prevalent we expect each label or model to be in the population. In a true empirical study this might be uniform—if we had no reason to suspect that a model may be more likely than another *a priori*, or we may have substantive reasons for weighting models. For example, models may capture the prevalence of a genetic factor, and we may have quantitative evidence for that prevalence. In a theoretical study such as this one, we are simulating equal numbers of samples from each theoretical model, and thus the term will be constant and cancel out. The data points in a classification problem are given, and thus the denominator is also a constant. The most probable class for a given data point reduces, therefore, to the mode of the likelihood function:

$$Y_{pred} = \arg\max_{y} \mathbb{P}(X_1, \ldots, X_p|Y) \tag{3.2}$$

This is the *Bayes classifier* for a controlled simulation experiment, and its error rate in separating data points by model is called the *Bayes error*. This is the lowest possible error in separating the models given the data (Devijver and Kittler, 1982; Fukunaga, 1990; Hastie et al., 2009). The Bayes error is zero when we can correctly identify each data point as to its model of origin (as in the left panel of Fig. 3.2, and rises as two models overlap in the measurement space. With sufficient overlap, the

Bayes error could approach 0.5, which represents a prediction rule which is no better than chance, or conceivably rise even further, indicating that our classifier performs even worse than coin-flipping.[1]

Unfortunately, we can almost never directly calculate the Bayes error rate for a prediction or classification rule, because we rarely have an expression for the likelihood function of our transmission models in the spaced formed by the predictor variables. Bayes error can be directly calculated, in fact, only for a small number of cases, such as Gaussian distributions with a shared covariance matrix. There is a large literature, especially in pattern recognition and language classification, on approximating upper bounds for the Bayes error of a classifier, because it is highly useful to know when you cannot improve a recognition system or classifier any further (Antos et al., 1999; Dobbin, 2009; McLachlan, 1975). Most such upper bounds are based upon parametric models, and use estimates of a distance metric between the classes being distinguished (typically, the Mahalanobis or Bhattacharyya distance) (Devijver and Kittler, 1982). Such bounds are difficult to apply in situations where we have complex social learning models, whose probability density functions in the space of measured variables are typically unknown.

Despite the fact that we can rarely calculate the Bayes error rate, it is useful as an operational definition for equifinality, since it measures our uncertainty about model choice given a set of measurable variables. In practice, we approximate the Bayes error by employing classifier algorithms which are able represent complex relationships between all of the predictor variables in order to get as close to the Bayes error rate as possible. This generally means using methods with higher model "capacity" than the linear models familiar to most archaeologists (such as logistic regression or linear discriminant analysis). Formally we seek methods with high Vapnik-Chervonenkis (or VC) dimension, which measures the ability to represent complex decision boundaries between classes (Vapnik, 2013). At present, methods such as boosting, bagging, and ensemble approaches

---

[1] Error rates can obviously be higher than 50% in an empirical study, but in a controlled simulation study like the present paper, error higher than a coin-flip is indicative of a problem with the experimental setup.

that combine weaker methods such as decision trees with boosting combine high capacity with very low error across many benchmark cases (Hastie et al., 2009), and thus come closest to estimating the Bayes rate (Tumer and Ghosh, 2003).

3.3.3 Simulation Modeling of Cultural Transmission Mixtures

In order to measure the overlap between various theoretical mixtures of social learning strategies, I built simulation models for the scenarios given in Table 3.1. The outcomes of all four transmission models are derived by simulating the dynamics of the model in an agent-based framework that allows each agent to be assigned a different transmission rule. All simulations employ the Moran dynamics, where one individual engages in a copying event at each elemental step (Moran, 1962, 1958; Aoki et al., 2011). Innovations are modeled using the "infinite alleles" approximation, where every innovation is new to the population (Ewens, 2004). Simulations were performed using the CTMixtures software package, available as open source software.[2] The parameters for all simulation runs are given in Table 3.2. Where there is a range given (e.g., innovation rate), the parameter is treated as a prior distribution and each simulation run is assigned a uniform random value from the range. This ensures good coverage of the parameter space given 25,000 replicates for each of the 4 models.[3]

Model Code	Model Description
Unbiased	Population with only unbiased copiers
Equal Mixture	Equal numbers of conformists and anti-conformists
Conformist Dom	70% conformists with 30% anti-conformists
AntiConf Dom	70% anti-conformists with 30% conformists

Table 3.1: Theoretical models that were simulated as part of this study.

---

[2] https://github.com/mmadsen/ctmixtures

[3] The use of a good prior distribution for parameter ranges also results in simulation data that are usable for later data fitting by approximate Bayesian inference (Beaumont, 2010; Crema et al., 2014; Csilléry et al., 2010; Marin et al., 2012).

Parameter	Value or Interval
Innovation rate (in θ scaled units)	$[0.1, 5.0]$
Probability of conformism	$[0.05, 0.25]$
Probability of anti-conformism	$[0.05, 0.25]$
Sample fractions	0.1 and 0.2
Time averaging intervals (units of 100 individuals)	10, 20, 50, 100
Population size	100
Number of trait dimensions (loci)	4
Initial traits per dimension	10

Table 3.2: Parameters for simulation runs across the four models studied. Intervals are treated as prior distributions, and each simulation run is assigned values derived from a uniform random sample on the interval indicated. Lists of values are all applied to every simulation run (e.g., there is both a 10% and a 20% sample from each simulation run. Single values are applied to every simulation run, and represent a point prior.)

Simulated populations are 100 individuals in size, because most archaeological studies of cultural transmission have focused upon situations where population sizes are assumed to be small. Each simulated individual carries 4 different traits at any time, which are treated as separate loci or dimensions. Trait frequencies are tracked on a per-locus basis, and combinations of loci are tracked in order to simulate archaeological "types" or classes which include multiple dimensions of variation.

Regardless of transmission model, social learning involves no interaction effects between loci in this study. The population is seeded with 10 randomly chosen traits at each locus as a starting configuration. The evolution of each simulated population proceeds for 4 million elemental steps, which is equivalent to about 40,000 copying events on average per individual. This value was chosen by performing simulations at 1 million time step intervals and verifying that the distribution of a key statistic (the number of traits per Loci) had stabilized. This occurred in most cases between 2 and 3 million steps, and in all cases between 3 and 4 million, so the last value was chosen.[4] At the end of

---

[4] The analysis underpinning this decision is available in the Github repository at `https://github.com/mmadsen/experiment-ctmixtures/analysis/verification`.

4 million simulation steps, a suite of variables are measured from each of the 25,000 replicates and stored for analysis.

3.3.4 Summary Statistic Selection

Since most previous work on identifying transmission mode from archaeological data employ single diagnostic variables, and begin to display equifinality under realistic data collection conditions, it is reasonable to examine whether using multiple variables will yield more discriminatory power in the same contexts. By representing the outcomes of transmission models in a higher dimensional space, it should be easier to find a decision boundary ("separating hyperplane") that correctly predicts the model which generated each data point, if such a boundary exists.

The predictor variables chosen in this study focus upon measures of richness and diversity, trait survival over time (Kandler and Shennan, 2013), and the Slatkin neutrality test (Slatkin, 1996, 1994). Each has been employed in the archaeological literature on identifying cultural transmission modes, or is a variant on such measures (e.g., IQV is a normalized version of Shannon entropy), and crucially, all are measurable in standard archaeological contexts using attribute or class frequency data. This additionally makes most of the variables applicable to the re-analysis of already published data, which is an important usage scenario in archaeological research.

For the locus-centric variables, each statistic was applied to each locus separately, and the mean, minimum, and maximum of the values obtained for each locus were recorded. I collect order statistics in addition to the mean value, since it is possible that minima and maxima might be a better discriminator between models than averages. In addition to the variables calculated upon each of the 4 loci, the traits at each locus were combined into a cross-tabulation of "classes" which simulates the process of archaeological classification. Each class represents a different combination of traits from the 4 loci, and very roughly simulates observing cultural variation through the lens of

Variable	Model Variable
Cross-Tabulated Class Richness (Class)	num_trait_configurations
Slatkin Exact (Class)	configuration_slatkin
Shannon Entropy (Class)	config_entropy
IQV Diversity (Class)	config_iqv
Neiman T_f (Class)	config_neiman_tf
Slatkin Exact (Max for Locus)	slatkin_locus_max
Slatkin Exact (Min for Locus)	slatkin_locus_min
Slatkin Exact (Mean for Locus)	slatkin_locus_mean
Shannon Entropy of Trait Frequencies (Min)	entropy_locus_max
Shannon Entropy of Trait Frequencies (Max)	entropy_locus_min
Shannon Entropy of Trait Frequencies (Mean)	entropy_locus_mean
IQV Diversity Index (Min)	iqv_locus_max
IQV Diversity Index (Max)	iqv_locus_min
IQV Diversity Index (Mean)	iqv_locus_mean
Trait Richness (Min)	richness_locus_max
Trait Richness (Max)	richness_locus_min
Trait Richness (Mean)	richness_locus_mean
Kandler-Shennan Trait Survival (Min)	kandler_locus_max
Kandler-Shennan Trait Survival (Max)	kandler_locus_min
Kandler-Shennan Trait Survival (Mean)	kandler_locus_mean
Neiman T_f (Min)	neiman_tf_locus_max
Neiman T_f (Max)	neiman_tf_locus_min
Neiman T_f (Mean)	neiman_tf_locus_mean

Table 3.3: Variables measured from each transmission model simulation sample. The parenthetical expression records whether the variable was calculated for cross-tabulations of all 4 loci (Class) or represent the order statistics from individual loci (Min/Mean/Max). The right column records the variable name used within R statistical models, for examining the relative importance of each variable in classifying observations.

a standard paradigmatic classification (Dunnell, 1971). The same variables are then measured as a function of the class counts.[5] This allows us to understand whether transmission models are better distinguished on a per-locus (dimension) basis or by operating on more complex classes that

---

[5] The sole exception is the Kandler-Shennan survival time, which is not measured here for the cross-tabulated classes. Understanding the quantitative behavior of this measure for multidimensional classes of traits is an important open research question, however.

combine several traits together. The full list of measured variables is given in Table 3.3.

As a final note on variable selection, in an exploratory analysis for this project, I tried to include the power law exponent as a summary statistic, given the important work by Bentley and colleagues (2004) and Mesoudi and Lycett (2009). Proper fitting of power law distributions to empirical data is more difficult than simply determining if a straight line describes the data after double logarithmic transformation; in particular testing if a data set displays a power law is difficult if there are a small number of counts and small sample size (Clauset et al., 2007). This may be difficult in many archaeological cases where the number of classes may be relatively small (on the order of ten rather than hundreds), and the total sample sizes can be relatively small. In their study, Mesoudi and Lycett (2009) use the cumulative number of adoptions of each trait over the entire time span of their simulations as the "frequency" used to calculate power law exponents.[6] This quantity is available in simulation but would unobservable in empirical data. Whether or not to include power law exponents as predictor variables will depend strongly on the data set in question. Given the experimental setup used in this study, power law exponents were difficult to calculate in a stable way. Thus, I did not include them in the remainder of this study, and their discriminatory power in combination with other variables thus remains an open question.

3.3.5 Data Collection Treatments

At the end of each simulation run, after the model has reached a quasi-stable equilibrium (measured as stability in per-locus trait richness), a series of samples are taken from the evolving population. These samples are taken in ways that correspond to various real-world data collection strategies. First, a census of the entire population is taken. This functions as a baseline for the "most complete" information we can use to identify transmission modes, and there are also conditions during ob-

---

[6] I confirmed this by inspection of the source code for their simulation model, which was provided by Alex Mesoudi.

servational studies or in laboratory experiments where census is possible. In archaeological studies, anything approximating a census is usually impossible, although Jonathan Scholnick's study of New England gravestones and their makers may approximate this quality of data collection (Scholnick, 2012). Second, the simulated population is sampled, at the 10% and 20% levels. Sampled data is ubiquitous in archaeological research, and although the issues involved in mapping artifact samples to their meaning for the underlying population of social learners is complex and unresolved, it is useful to determine whether the overall sample fraction has a measurable effect upon model equifinality.

Archaeology derives its empirical data by sampling a sedimentary record of artifact discard and trace fossil creation by many individuals, often over large spans of time (Schiffer, 1983, 1987; Stein, 2001, 1987, 1993, 2001; Stein et al., 2003). Thus, our data almost never represent synchronic or "point in time" samples of the results of human activity (Grayson and Delpech, 1998; Lyman, 2003; Madsen, 2012; Porčić, 2014; Premo, 2014). Therefore, the simulated samples of cultural transmission in this study are also temporally aggregated over a number of time steps, and the aggregate trait counts used to determine the frequencies of cultural traits over the entire interval. The population census has no temporal aggregation, and thus does represent a synchronic census.

Time averaging is implemented according to the schematic in Fig. 3.3. At the end of the simulation run, sampling begins at a time index calculated to allow time averaged samples to be taken twice, with a gap of 50 "generations" to also allow the calculation of the Kandler-Shennan trait survival statistic (although unlike their original study, the values at the start and end times are inherently time averaged in this study, which would be the case in any real archaeological context) (Kandler and Shennan, 2013).

The data collection strategies employed in this study are given in Table 3.4. Applied to all 23 variables, the study yielded approximately 900,000 samples from the four transmission models listed

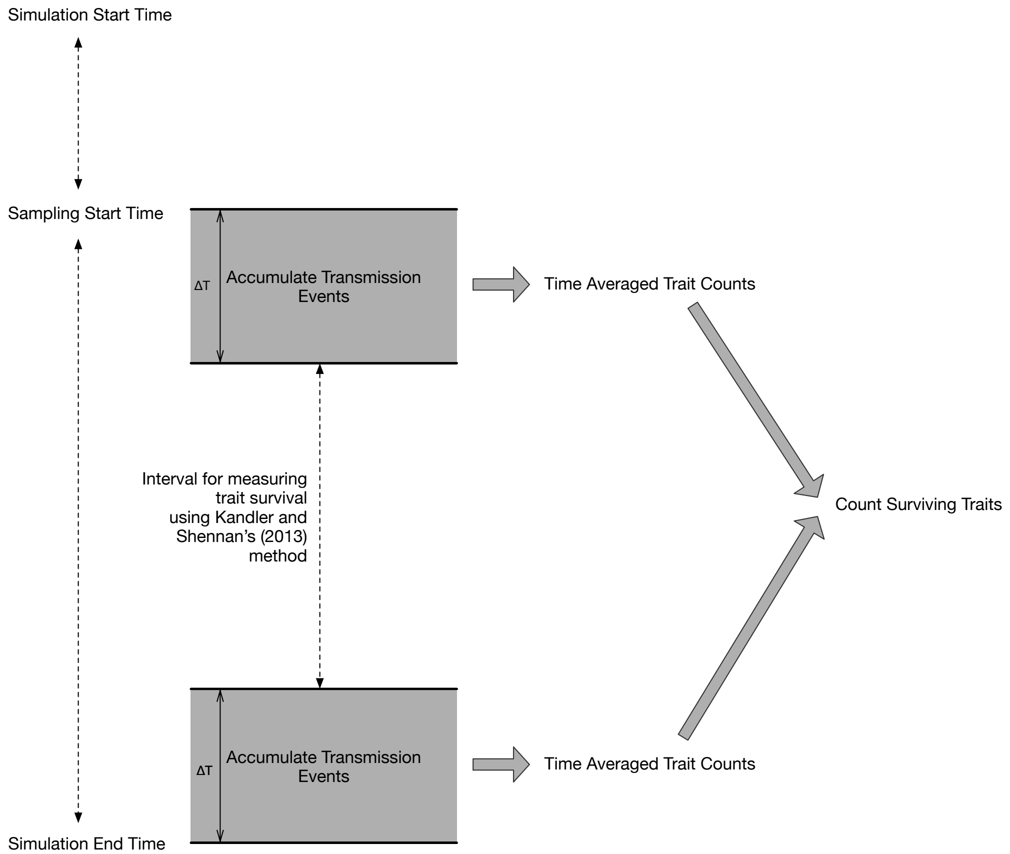


Figure 3.3: Schematic of how trait survival as described by Kandler and Shennan (Kandler and Shennan, 2013) is extended to time averaged samples of transmission events. Time runs from the start of the simulation run at the top, to the end at the bottom. The interval of time over which we calculate the Kandler-Shennan trait survival is given as a simulation parameter, and represents the gap in the middle of the diagram. Before and after that gap are sampling windows during which transmission events are accumulated over some number of simulated "generations" (values of 10, 25, 50, and 100 are used in this paper). Trait survival is then calculated as the number of traits present in the starting time averaged sample of transmission events, which are still present in the ending time averaged sample of events.

Sampling Strategy	Time Averaging Duration
Population Census	0
10% Sample	10
10% Sample	25
10% Sample	50
10% Sample	100
20% Sample	10
20% Sample	25
20% Sample	50
20% Sample	100

Table 3.4: Data collection strategies, applied to every simulation run. Time averaging duration is given in units of ”generations,” which are units of 100 time steps (given the population size). 100 generations thus represents 10,000 elemental time steps in the Moran simulation dynamics.

in Table 3.1.[7]

3.3.6 Classifier Selection and Training

Classifier algorithms are supervising learning models from statistics and machine learning that predict a categorical response from a mixture of discrete or continuous variables (Hastie et al., 2009). The most familiar classifiers in archaeological practice are logistic regression and discriminant function analysis, but neither is competitive with contemporary “ensemble” methods which combine many classifier rules into a single prediction. In such models, combining predictors can both reduce the variance of prediction (e.g., bagging added to traditional classifiers and random forests), and reduce bias. Some classifiers, like boosted trees, can do both.

Since the Bayes error rate of comparing two complex transmission models is not something we can calculate or even estimate, we must approximate it using the best performing classifier model available. A very general result in statistical decision theory (called, appropriately, the “No Free

---

[7] All data and analyses for this study are available as part of a Github repository, although large data files are kept on Amazon S3 for long-term storage. See `https://github.com/mmadsen/experiment-ctmixtures` for details. The published analysis described here is the “equifinality-4” data set.

Lunch" theorems) guarantee that there is no single prediction model that can achieve the best result with every data set and problem (Wolpert, 2002; Wolpert and Macready, 1997). Thus, I took a compromise approach, selecting several algorithms that are known to have excellent performance across a range of data sets, and then performing a pilot study using the four transmission models previously described. A recent study compared 179 classifier algorithms on 121 different data sets (representing the entire UC Irvine Machine Learning Database), and found that random forests (Breiman, 2001), support vector machines, and gradient boosted classifiers performed the best (Hastie et al., 2009). Additionally, some ensemble methods (random forests and gradient boosted classifiers) provide information on variable importance as an integral part of the algorithm. Since understanding which of our 23 variables are useful for separating transmission models is an important aspect of this study, I evaluated random forests against gradient boosted classification trees using small simulated samples from each transmission model.[8] Gradient boosted models outperformed random forests on these simulated data, are comparable in computational costs, and are used for all further results in this paper.

Gradient boosted classification operates by repeatedly fitting a set of decision trees to the data (Natekin and Knoll, 2013; Hastie et al., 2009). In each round, decision trees are fit to the training data, and individual data points scored as errors or successful predictions. Subsequent trees are fitted by modifying the trees in the direction that minimizes the residual error. This is equivalent to finding the gradient of the loss function in the space of possible classifier functions, hence the name of the method. The impact of each gradient step is smoothed by including a "shrinkage" factor. Finally, the gradient steps are "boosted" to weight data points by the success in prediction, such that data points that are frequently misclassified become targeted by the algorithm until they can be correctly predicted (Freund, 1995; Freund et al., 1999; Schapire and Freund, 2012). After a specified number

---

[8] The data for this initial comparison are available in the `https://github.com/mmadsen/experiment-ctmixtures` repository under the experiment name "equifinality-2".

of iterations, the class or label membership of each data point is obtained by having each gradient step classifier tree "vote" for class membership, and the final answer is the majority vote. This class of models can also be visualized as repeated refitting of residuals until error is minimized (Friedman, 2001). This combination of boosting and iterative function search is very powerful, and gradient boosted models regularly achieve top accuracy in benchmark studies.

In this study, I employ the R package (**gbm**) for gradient boosted classification (Ridgeway, 1999), with the binomial deviance $\log(1+\exp(-2y\hat{y}))$ as our loss function, where y is the true model for a data point, and $\hat{y}$ is the classifier model's prediction. Binomial deviance approximates the "zero-one" loss function with one which is differentiable, which is needed for a gradient descent method. The tuning parameters for this study (number of boosting iterations, depth of classification trees) were selected using 5 rounds of repeated 10-fold cross-validation on the training data (Kim, 2009; Kuhn and Johnson, 2013).

The full data set of simulation samples, after data collection treatments, was split into two chunks. 80% of the data were used to train the classifier model, and 20% were held back to provide an unbiased evaluation of classifier performance. For each comparison of models reported, the training data were thus fitted 50 times across different values of the tuning parameters (number of boosting iterations, and depth of decision trees), and the best performing parameters chosen from the repeated cross-validation sets. The final model is then constructed using the entire training set and the optimal parameter values. All classifier tuning, final model fitting, and test error evaluation was performed using Max Kuhn's superb **caret** package for R (Kuhn, 2008; Kuhn and Johnson, 2013). The final results reported are those achieved on the 20% of data points held out for evaluation.

3.3.7 Classification Error and Equifinality Assessment

The basic data for assessing the quality of a classifier model is the *confusion matrix*, which compares classification successes and errors for a data set. A hypothetical example is given in Table 3.5. The most basic measure of classification quality is the *accuracy*, or the ratio of correct predictions to the total number of data points. In the confusion matrix, this is the ratio of the sum of diagonal elements to the sum of off-diagonal elements. In the example given in Table 3.5, the classifier is 82.5% accurate. We often also use the misclassification rate, which is simply $1 - accuracy$.

	Actual Model:	
Predicted	Model 1	Model 2
Model 1	**9000**	2500
Model 2	1000	**7500**

Table 3.5: Example confusion matrix. Columns correspond to the actual model for data points, rows correspond to predictions from a classification model. Bold numbers on the diagonal correspond to correct predictions, the off diagonal elements correspond to classification errors.

When the classes being predicted are not balanced, and especially if there are a small number of one class compared to another, a better statistic is Cohen's "kappa" (Kuhn and Johnson, 2013), which compares observed accuracy to what one would expect purely from chance, given the marginal totals:

$$\kappa = \frac{O - E}{1 - E} \tag{3.3}$$

where O is the observed accuracy, and E is the expected accuracy due to chance given the ratio of classes in the marginal totals of the confusion matrix. Kappa ranges from -1 to $+1$, with 0 indicating no agreement between predictions and the real class memberships. High values indicate good agreement, while values below 0.5 and especially less than 0.2 indicate very poor predictive ability (Altman, 1991). In the present context, a classifier comparison (for example, biased versus neutral models with no sampling or time averaging) that yield a high kappa value are strong evidence that no equifinality exists between the two situations, since the classifier is highly accurate. Low kappa values are evidence that despite strong statistical methods and many variables to choose from, we cannot distinguish between models, and thus the models may be equifinal.

3.4 Results

From the four models listed in Table 3.1, I examine three comparisons. The first comparison addresses our general capability to distinguished unbiased (or neutral) cultural transmission from samples drawn from populations with mixtures of both conformist and anti-conformist propensities. The second examines our ability to distinguish between an unbiased population, and a population comprised of a majority of conformists, with 30% anti-conformists, to create a mixture of social learning modes. The third compares an unbiased population with a majority of anti-conformists, with 30% conformists, to produce the opposite configuration.

For each of the three comparisons, the Cohen's kappa scores were calculated two ways: with all predictive variables (including the class variables measured on intersections of attributes from the 4 different loci), and for classifiers built only with per-locus variables (i.e., where we look only at what would be single attributes rather than more complex archaeological classes or culture-historical types.

3.4.1 Classification Error and Equifinality Results

In this section I outline the results from the three model comparisons described above, across all 9 data collection treatments. Figure 3.4 depicts Cohen's kappa results for the classifier trained with all predictor variables, including those from per-locus trait counts, and those from intersecting all four loci into "classes" that mimic the way paradigmatic classifications are constructed to build archaeo-

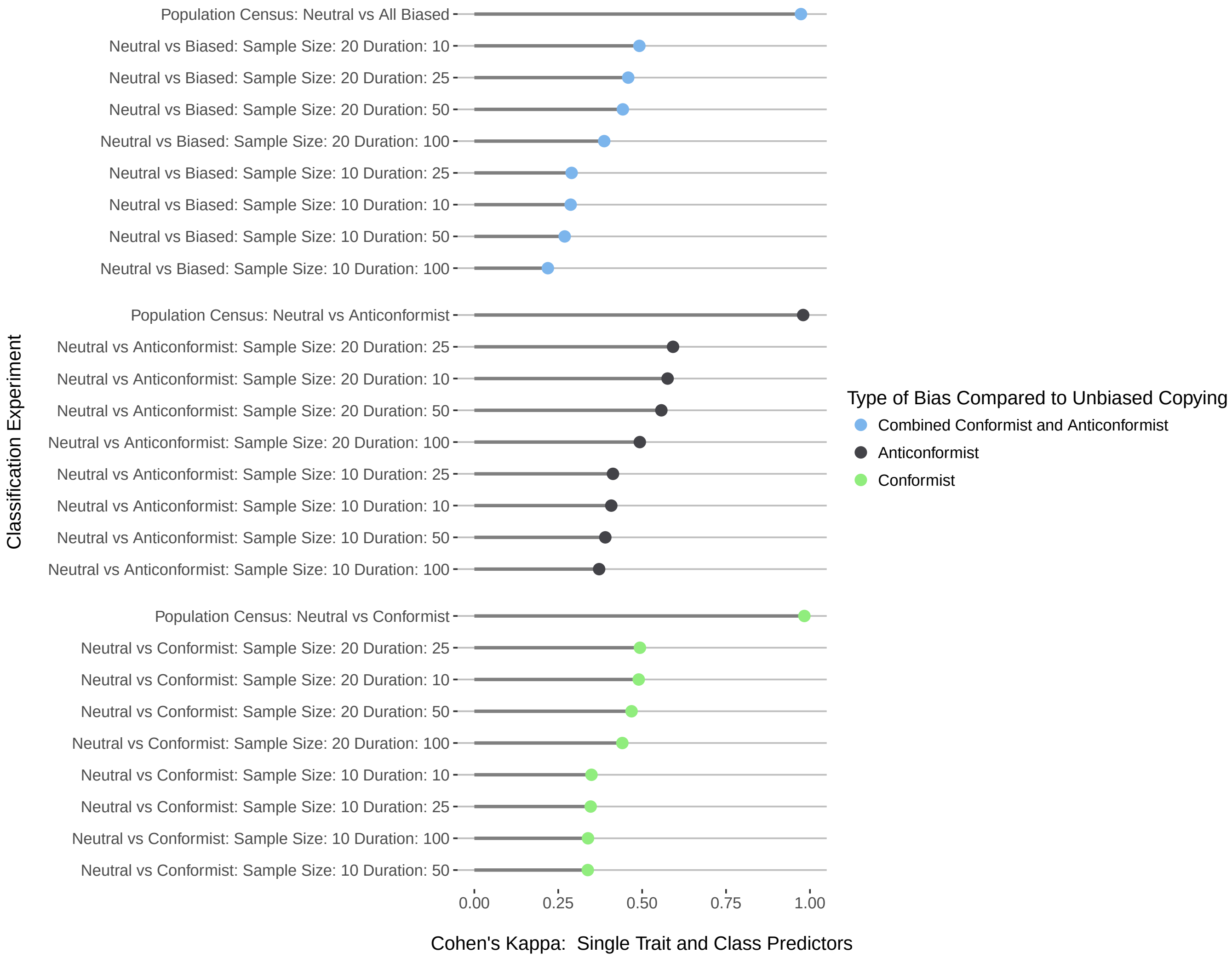


Figure 3.4: Cohen's kappa values for model comparisons, using a classifier trained on all predictors including per-locus and "classes" built from intersecting loci and then counting frequencies. Top panel provides results for comparing a neutral, unbiased population and a balanced mixture of conformist and anti-conformists. The middle panel provides results for unbiased compared to a population dominated by anti-conformists. The bottom panel depicts results for unbiased compared toa population dominated by conformists. Each panel presents results for the 9 data collection treatments analyzed.

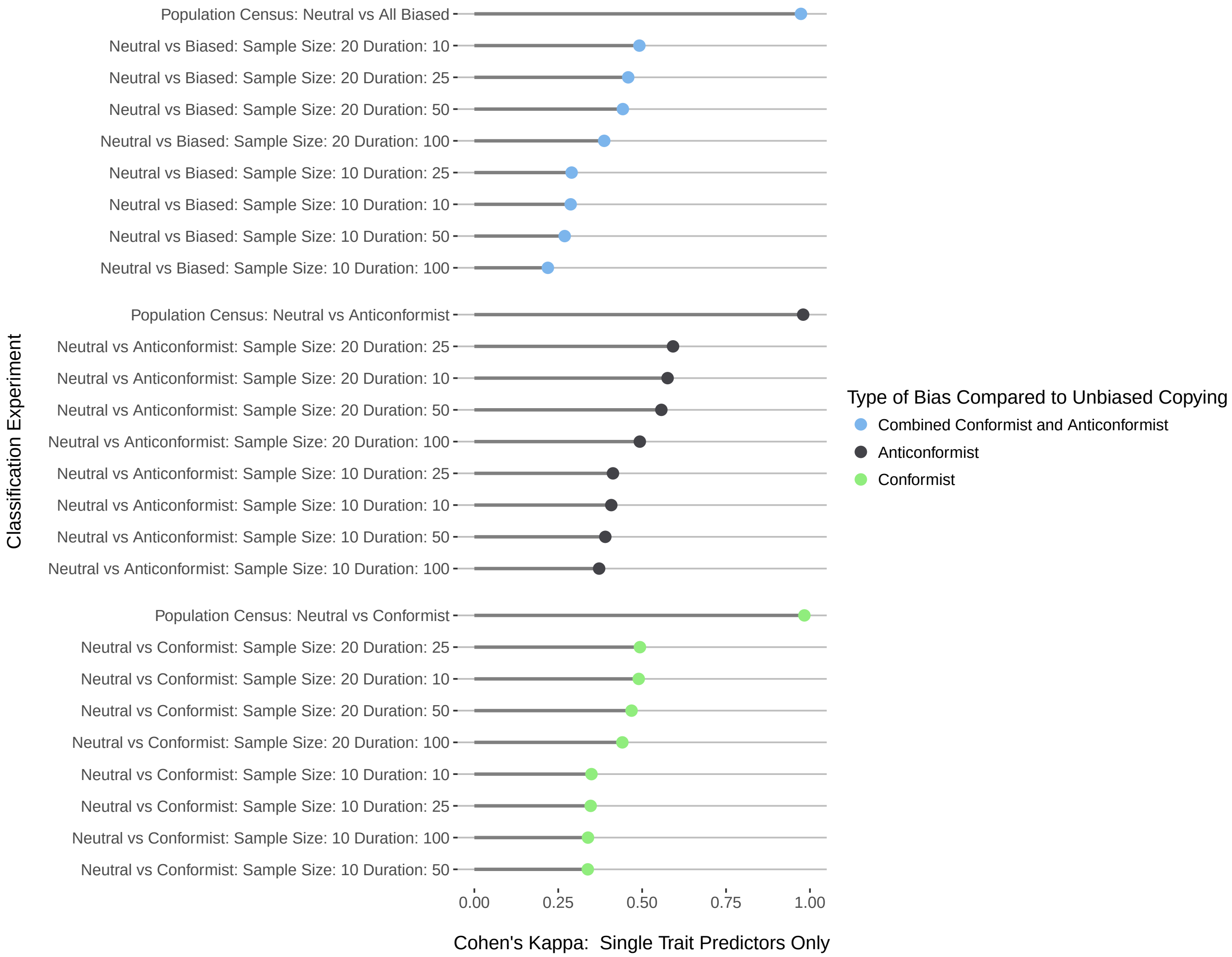


Figure 3.5: Cohen's kappa values for model comparisons, using a classifier trained only on predictor variables derived from per-locus trait counts. Top panel provides results for comparing a neutral, unbiased population and a balanced mixture of conformist and anti-conformists. The middle panel provides results for unbiased compared to a population dominated by anti-conformists. The bottom panel depicts results for unbiased compared toa population dominated by conformists. Each panel presents results for the 9 data collection treatments analyzed.

logical types (Dunnell, 1971). Figure 3.5 provides the Cohen's kappa results for the classifier trained only on predictor values derived from per-locus trait counts.

What is strongly apparent, however, is population-level, coarse grained data *can* reliably distinguish between a population of unbiased copiers and various mixtures of individuals with differing propensities for conformism and anti-conformism. Using multiple variables in a classifier model creates sufficient discriminatory power that we can distinguish these models reliably, even in a balanced mixture with equal numbers of conformists and anti-conformists. The effects do not, apparently, "cancel out."

This ability to distinguish models, however, is only apparent where we can census the entire population (or, presumably take very large samples), and where there is no time averaging. We should, in other words, expect to be able to identify pure and mixed models of social learning modes in situations where we can obtain synchronic samples, and capture enough of the population that we can accurately capture all of the variation present. This is possible in experimental contexts, and in many contemporary data sets that represent data on whole populations. That's good news for researchers seeking to use microevolutionary models to study social psychological phenomena related to social learning in contemporary contexts.

Even more strongly apparent, however, is that accuracy rapidly declines as sample fraction decreases and as time averaging increases. As one might expect, larger samples offer more accurate predictions than smaller samples. Within the larger, 20% sample, when cross-tabulated class and per-locus predictors are included, accuracy is highest with the smallest amount of time averaging (10 generations), and decreases as time averaging increases. When we remove cross-tabulated class predictors, and simply look at per-locus variables, this clean pattern is not apparent, and accuracy is not a function of time averaging duration. Furthermore, for the smaller 10% sample with all variables included, accuracy is not a function of time averaging duration. In these cases, Cohen's kappa values are between 0.2 and 0.3, indicative of a very poor classification model with very little predictive power.

While sample size is partially addressable in data collection, the degree to which archaeological deposits are aggregated over some window of time is often not alterable in research design. There is equifinality here which we simply cannot avoid in most archaeological cases, which renders our ability to distinguish between realistic mixtures of social learning modes and unbiased copying unreliable at best, and impossible at worst.

3.4.2 Which Predictor Variables Help Discriminate Models?

Since there is some predictive power here, particularly using population census information, it is important to understand which variables were providing that discriminatory power. Gradient boosting algorithms allow measurement of how much each predictor variable contributes to a classification model. The importance of a variable is assessed over the iterations of tree construction by estimating the relative improvement in training set misclassification error from adding the variable to the model. The importance values are usually scaled such that the most important variable has a score of 100, and variables with smaller importance values are less important to classification power.

In general the pattern is the same across comparisons, so I illustrate variable importance by examining the comparison between unbiased copiers, and the population composed of a balanced mixture of conformists and anti-conformists. Table 3.6 gives the relative importance of predictor variables when the classifier includes classes created by intersecting the 4 loci with individual traits.

Most classificatory power in this comparison comes from the richness in the cross-tabulated classes. In general, richness (whether class or per-locus) dominated all of the classifier models, which is not surprising given the central role of the amount of variation in mathematical treatments of the Wright-Fisher and Moran models (Ewens, 2004). The second variable in importance in this comparison is the p-value for a Slatkin's "exact" test, which determines the probability that a given set of class frequencies come from the Ewens Sampling Distribution (Slatkin, 1994). The Shannon entropy, a measure of evenness in frequency (in this case averaged across loci) provides about half as much predictive power as richness. Interestingly the Kandler-Shennan survival time, averaged across the 4 loci, offers significant discriminatory power, suggesting that the method used here for simulating it along with time averaging is a fruitful avenue for additional research. Beyond this, the predictors have increasingly little power, likely because we have multiple variants on evenness measures in the full list of summary statistics, and their results are collinear with other variables.

Importance	Predictor Variable
100.00	Cross-Tabulated Class Richness
85.49	Slatkin Exact for Classes
47.12	Shannon Entropy (Mean for Locus)
24.12	Kandler-Shennan Trait Survival (Mean for Locus)
18.88	IQV Diversity (Mean for Locus)
10.85	Shannon Entropy for Classes

Table 3.6: Relative importance of predictor variables for population census data, in the comparison between unbiased transmission and a balanced mixture of pro- and anti-conformists. The most important variable is (by convention) scaled to 100, and the values indicate the ratio of variable importance to the variable which is most effective at classifying data points. Only values greater than 10 are shown. The remainder of the predictor variables are 1/100th as effective as class richness or less.

3.4.3 Time Averaging Makes Identification of Bias More Likely

As shown in these results and previous studies (Madsen, 2012; Porčić, 2014; Premo, 2014), the temporal aggregation we encounter in archaeological data is a significant barrier to discriminating between theoretical models of cultural transmission. Equifinality increases as the amount of time averaging represented in a data set increases. But this is not the whole story. Time averaging, combined with sampling effects, do not simply make our classifiers worse in a random manner. Instead, there is substantial bias in the pattern of classification (and thus model identification) errors.

We can see this clearly by examining two confusion matrices from the comparison between unbiased and a balanced mixture of social learning biases. Table 3.7 shows the pattern of correct and incorrect predictions for the population census with no time averaging treatment. In the confusion matrix, the rows depict the *predicted* data generating model, and the columns provide the true data generating model. Thus, in the row "biased", the first column are correct predictions that data points came from the mixture of biases model, while the second column are incorrect predictions that the data points arose from a model of bias, when the data points in fact arose from unbiased copying.

Even in the population census treatment, there is a slight tendency for prediction errors to be unbalanced. It is more likely to misidentify a data point arising from unbiased copying as coming from a mixture of biased copiers than it is to make the opposite error. In other words, we over-identify bias, even with complete data, by about 30

Table 3.8 shows the pattern of errors for the same comparison, but with 20% sample fraction and time averaging duration of 50 steps. This is not the worst treatment for accuracy by far, but note that the pattern of errors has become significantly more asymmetric. With this data collection treatment 2.7 times more likely that we will make an error favoring an identification as transmission bias than that we will misidentify bias as neutrality.

By looking at the ratio of the right column in each confusion matrix, across all data collection treatments, we can see how the magnitude of this asymmetric preference for identification of bias scales with sample size and duration of time averaging (Table 3.9). The smaller our samples are from the original population, and the more temporal aggregation in our assemblages, the more likely we are to "see" biased cultural transmission, even with combinations of summary statistics and powerful classifier models.

3.5 Discussion

The central aim in this study has been to examine whether realistic mixtures of social learning strategies present fundamental equifinalities that were impossible to distinguish using population level summary data. The results demonstrate that mixtures of biased social learning strategies do not "cancel out" sufficiently to mask their statistical signatures—when you have enough data, synchronic observations, and use several summary statistics in combination.

However, when time averaging is present, and especially with small sample sizes, our ability to discriminate between theoretical models essentially disappears. Equifinality is omnipresent. Even worse, the results indicate that our ability to distinguish between models does not degrade symmetrically. In the presence of time averaging and with small samples, we are several times more likely to identify data points as arising from transmission bias than we are when we have synchronic and more complete data.

Archaeologists engaged in the study of microevolutionary cultural transmission models should employ a healthy skepticism about the enterprise of fitting artifact class frequency data to models. Even with diachronic methods such as trait survival (Kandler and Shennan, 2013), multiple summary statistics, and powerful analytic methods, there is little evidence that we can distinguish these models with simulated data. And that should give us great pause in thinking that our conclusions from analyzing archaeological samples are sound. It is time to evaluate whether the application of microevolutionary models in archaeology is a fruitful enterprise, or whether our time is better spent deriving models which are robust in the face of time averaging—even if those models address larger scale questions instead of the social psychological questions involved in most dual inheritance theory.

It is certainly true that evolutionary change is ultimately caused by individual variation and how that variation affects success. In a sense, all evolutionary change is "caused" by that individual level variation; patterns and processes we observe at larger spatiotemporal scales and taxonomic levels are "upper level effects" of individual level variation (Walsh, 2019). This does not make population, regional, and taxonomic level patterning in evolutionary history mere "epiphenomena." Far from it. The structure and history of populations and their environments provide the fitness landscapes which shape and sort variation among individuals. Archaeology is uniquely placed not just to document variation at larger scales, but to study processes that are only operative at scales larger than individuals and single populations.

We do not have a complete catalog of evolutionary processes that can only be studied above the level of single populations, but there are numerous examples. Variation in traits which affect dispersal lead to phenotypic variation among populations depending upon their migration and dispersal characteristics, an effect that Shine et al. (2011) refer to as "spatial sorting." We can and should expect such differences to be evidence at archaeological scales when we compare sedentary versus mobile communities in the archaeological record, for example. Niche construction often occurs over evolutionary time (Laland et al., 2000; Odling-Smee et al., 2003; Odling-Smee, 2007), including all kinds of coevolutionary relationships such as domestication (Rindos, 1984). All of these processes require the time depth inherent in the archaeological record for proper study and explanation. Large scale population structure—patterns in interaction between communities, is something we can document at mesoscopic and macroevolutionary scales, and is part of what we mean when we discuss "social complexity." Building models for these kinds of processes requires tools for creating diachronic models and fitting them to coarse grained, time averaged data. In the following chapters, I turn to several methods for creating such models and evaluating equifinality among them.

	biased	neutral
biased	14898	132
neutral	102	4868

Table 3.7: Confusion matrix for comparison of unbiased versus mixed bias models, for the data treatment with population census of trait frequencies and no time averaging.

	biased	neutral
biased	13926	2724
neutral	1074	2276

Table 3.8: Confusion matrix for comparison of unbiased versus mixed bias models, for the data treatment with 20% sampling fraction and 50 steps for time averaging duration.

Data Collection Treatment	% of Unbiased Data Points Misclassified as Biased
Population Census	2.6
Per-Locus Population Census	3.4
Sample Size: 20 Duration: 10	49.1
Sample Size: 20 Duration: 25	49.5
Per-Locus Sample Size: 20 Duration: 25	49.7
Per-Locus Sample Size: 20 Duration: 10	50.1
Per-Locus Sample Size: 20 Duration: 50	54.4
Sample Size: 20 Duration: 50	54.5
Sample Size: 20 Duration: 100	57.6
Per-Locus Sample Size: 20 Duration: 100	58.1
All Sample Sizes and TA Durations	68.0
Per-Locus Sample Size: 10 Duration: 25	73.5
Sample Size: 10 Duration: 25	73.6
Sample Size: 10 Duration: 50	73.7
Per-Locus Sample Size: 10 Duration: 10	73.8
Per-Locus Sample Size: 10 Duration: 100	74.0
Per-Locus Sample Size: 10 Duration: 50	74.1
Sample Size: 10 Duration: 100	74.4
Sample Size: 10 Duration: 10	74.7

Table 3.9: Percentage of data points from the unbiased transmission model that are falsely identified as arising from a biased model.

CHAPTER 4

Combinatorial Structure of the Deterministic Seriation Method with Multiple Subset Solutions

Abstract Seriation methods order a set of descriptions given some criterion (e.g., unimodality or minimum distance between similarity scores). Seriation is thus inherently a problem of finding the optimal solution among a set of permutations of objects. In this short technical note, we review the combinatorial structure of the classical seriation problem, which seeks a single solution out of a set of objects. We then extend those results to the iterative frequency seriation approach introduced by Lipo et al. (1997), which finds optimal subsets of objects which each satisfy the unimodality criterion within each subset. The number of possible solutions across multiple solution subsets is larger than n!, which underscores the need to find new algorithms and heuristics to assist in the deterministic frequency seriation problem.

Source Posted to Arxiv.org (`https://arxiv.org/abs/1412.6060`), in December 2014. Co-authored with Carl P. Lipo.

4.1 Single Seriation Combinatorics

Seriation, whether employing class frequencies or simple occurrence to order assemblages, yields solutions which are permutations of the set of assemblages. Because we cannot determine the "polarity" of a seriation solution—which ends represent early and late—from the class data alone, a unique seriation solution is thus formally a pair of mirror-image permutations:

$$\{a, d, b, c, e\} \equiv \{e, c, b, d, a\} \tag{4.1}$$

This means that a set of n assemblages can yield $n!/2$ distinct solutions, regardless of whether solutions are composed of ordered similarity matrices or"Fordian" frequency curves. With small numbers of assemblages, enumeration and testing of all possible solutions is easy, even without parallel testing across many processors. The ability to test solutions by enumeration quickly breaks down with only a modest number of assemblages. Table 4.1 gives the number of unique solutions for selected problem sizes between 4 and 100 assemblages, and estimates of processing time to enumerate and test all solutions, assuming a cluster of 64 cores, and 5×10^{-4} seconds per solution test.[1] With 10 assemblages, we can test all solutions quickly enough that even a serial algorithm on a single core will be adequate to find the global best solution in a matter of hours, with parallelism improving this to real time responses.

A typical characteristic of many combinatorial algorithms is that small changes in problem size can have massive changes in processing time. 13 assemblages will turn out to be the practical limit for direct enumeration, even given parallel processing with circa-2012 technology, with total processing time of nearly 3 days running 64 cores at full capacity.[2] Problems involving 14 and 15 assemblages

---

[1] These assumptions concerning per-trial processing time and parallelism are arbitrary but within reach of social scientists given Amazon's EC2 cloud computing infrastructure, without requiring formal "supercomputer" access. Modification by a factor of 10 has little effect on the results, perhaps shifting feasibility upward slightly before combinatorial explosion occurs.

[2] Realistically, almost nobody would contemplate doing this, given the expense of the computing time relative to the value of guaranteeing the optimal solution, but the hypothetical example demonstrates that such solutions are *feasible*.

N	Seriation Solutions	Seconds	Years
4	12	9.4e-05	3e-12
6	3.6e+02	0.0028	8.9e-11
8	2e+04	0.16	5e-09
10	1.8e+06	14	4.5e-07
12	2.4e+08	1.9e+03	5.9e-05
13	3.1e+09	2.4e+04	0.00077
14	4.4e+10	3.4e+05	0.011
15	6.5e+11	5.1e+06	0.16
16	1e+13	8.2e+07	2.6
20	1.2e+18	9.5e+12	3e+05
40	4.1e+47	3.2e+42	1e+35
60	4.2e+81	3.3e+76	1e+69
80	3.6e+118	2.8e+113	8.9e+105
100	4.7e+157	3.6e+152	1.2e+145

Table 4.1: Number of unique seriation solutions and parallel processing time for sets of assemblages $4 < n < 100$, testing solutions across 64 cores, assuming 5ms per trial

reach the point where large clusters require more than a month and 19 months respectively, to solve. Beyond 15 assemblages, a “combinatorial explosion” sets in, with 20 assemblages requiring more than 3 million years, before solution times quickly exceed the lifetime of the universe.

In short, top-down enumerative methods are feasible for small sets of assemblages, and given widespread availability of multiple core computers, seriation packages should employ enumeration for small problems, or to build and test smaller parts of larger seriation solutions.

4.2 Deterministic Seriation with Multiple Solution Groups

In an earlier paper (Lipo et al., 1997), we introduced an iterative method for finding deterministic solutions to the frequency seriation problem by partitioning assemblages into subsets, each of which meets the unimodal ordering principle, within tolerance limits governed by sample size. Lipo (2001b) extended and refined the method in his dissertation research. Our initial work on

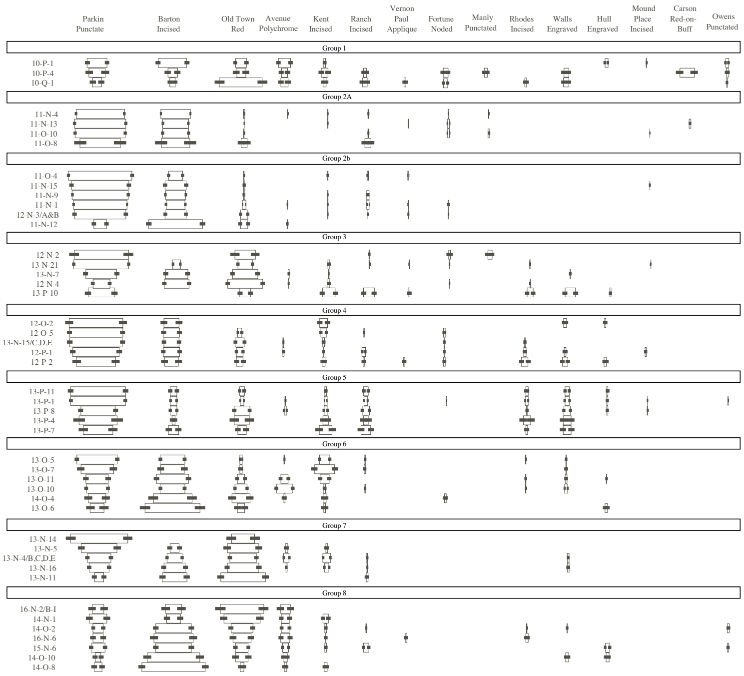


Figure 4.1: Example of a deterministic frequency seriation with assemblages partitioned into multiple subsets or solution groups. From Lipo (2001b), Figure 4.4.

the method employed a combination of automated calculations (e.g., bootstrap significance tests for pairwise orderings), and manual sorting of assemblages into groups and specific positions (using an Excel macro package available at `http://lipolab.org/seriation.html`). Figure 4.1 is an example of seriation with multiple solution groups, from Lipo's dissertation research in the Lower Mississippi Valley.

Our initial work suggests assemblages seriate together into groups reflecting variation in the intensity of cultural transmission among assemblages, over their duration of accumulation. In most cases, solution groups tend to be spatiotemporally compact, and form clusters when mapped on the landscape, although long-distance connections between past communities can also yield patterns which are more complex and less cohensive when mapped. Madsen's dissertation research is aimed at tying the properties seriation solution groups to their causes in regional patterns of interaction and the dynamics of specific cultural transmission models.

# of Solution Groups (m)	20	40	60
3	5.8e+08	2e+18	7.1e+27
4	4.5e+10	5e+22	5.5e+34
6	4.3e+12	1.8e+28	6.8e+43
8	1.5e+13	3.2e+31	3.8e+49
10	5.9e+12	2.4e+33	2.7e+53
15		2.9e+34	2.2e+58
20		1.6e+32	1.7e+59
25			3.7e+57
30			9.6e+53

Table 4.2: Number of ways to form m subsets (seriation solutions) from 20, 40, and 60 assemblages

In this section, the goal is to understand the complexity of the multiple seriation groups problem, constructing reasonable upper bounds for a given problem size, even if some problems encountered in real analyses do not approach the worst case. From a combinatorial standpoint, seriation with multiple solution groups has the following structure. We begin with n assemblages in total, and seek a solution or solutions whereby we end up with m solution groups, where $m < n$. Each solution must have at least one assemblage, and in practice will often have 3 or more (singletons may indicate assemblages which simply do not "fit" with anything else in the data set). The number of ways that n objects can be partitioned into m non-empty subsets (or solution groups) is given by the Stirling numbers of the second kind, which are given by the recursion equation:

$$\left\{ {n \atop m} \right\} = m \left\{ {n-1 \atop m} \right\} + \left\{ {n-1 \atop m-1} \right\} \tag{4.2}$$

Table 4.2 gives the number of ways to form a specific number of subsets (or seriation solution groups) from sets of assemblages ranging from 20 to 60. Each column runs from 3 solution groups to half of the number of assemblages, since the number of possible subsets is maximized just before $n/2$ and declines thereafter (Figure 4.2).

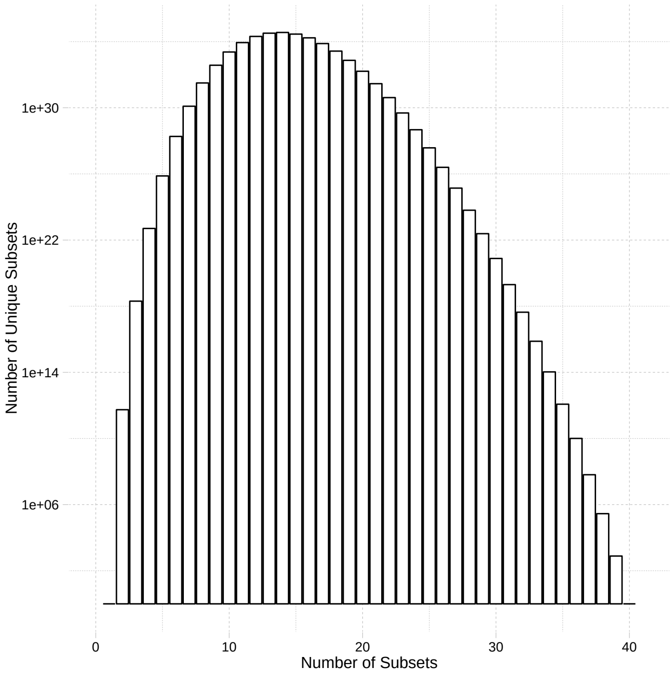


Figure 4.2: Number of Unique Solution Sets for 40 Assemblages When Partitioned Into m Solution Groups

We can immediately see that there are an enormous number of possible subsets for any assemblage size. There are fewer subsets, of course, than complete permutations of the set of assemblages since subsets are unordered (i.e., $\left\{ n \atop m \right\} < n!$ for all m). However, in the multiple seriation group problem, the problem size is larger than the corresponding Stirling number because we do not know in advance how many groups (subsets) a set of assemblages will seriate into. Thus, the total number of unique subsets which might contain the optimal solution is the total of the number of subsets, across all subset sizes:

$$\sum_{i=1}^{n} \left\{ n \atop i \right\} \tag{4.3}$$

This result is still smaller than the total permutations for a set of n assemblages. For example, given 40 assemblages, $n! = 8.159 \times 10^{47}$, whereas the total from Equation 4.3 for 40 assemblages is 1.575×10^{35}.

Another factor to consider is that each of these unique subsets resulting from a partition of n assemblages into seriation groups is still unordered. For example, if we partition 10 assemblages into 3 solution groups, there are 9330 unique ways of assigning the 10 assemblages to the 3 solution groups. Each group within a partition will have n_i members, where $\sum n_i = n$. The number of unique seriations for each of the 3 solution groups is $n_i!/2$, but we cannot assume that solution groups will have a balanced or equal number of assemblages (as Figure 4.1 does). Partitions such as:

$$\{1,2,3,4,5,6\}\{7,8\}\{9,10\}$$

are common in seriating real assemblages (Lipo, 2001b).

Since the factorial function grows so quickly, the computational cost of determining the correct permutation within a given seriation solution group is controlled by the size of the largest subset, especially if the other subsets are relatively small, as in the previous example. At worst, for a solution set with m solution groups, $m-1$ solution groups will contain 1 assemblage each, and the last solution group will consist of the remaining $n-m-1$ assemblages. This means, of course, that the worst case would involve consideration of on the order of $(n-m-1)!$ permutations within each solution group, for each of the subsets given by Equation 4.3. This yields:

$$\sum_{m=1}^{n} \left\{ {n \atop m} \right\} (n-m-1)! \tag{4.4}$$

Table 4.3 gives the total number of possible solutions for assemblages ranging from 4 to 100, where solutions may fall into multiple seriation groups of any size.

4.3 Discussion

Clearly, in the worst case, the combinatorial complexity of the multiple seriation groups problem is much worse than even the straight factorial case involved in single solution permutations. The

N	Total Solutions	Seconds	Years
4	15	0.00012	3.7e-12
6	4.7e+02	0.0037	1.2e-10
8	5.2e+04	0.4	1.3e-08
10	1.5e+07	1.1e+02	3.6e-06
12	8.5e+09	6.6e+04	0.0021
13	2.6e+11	2e+06	0.064
14	8.9e+12	7e+07	2.2
15	3.5e+14	2.8e+09	87
16	1.6e+16	1.2e+11	3.9e+03
20	1.7e+23	1.3e+18	4.2e+10
40	9e+65	7e+60	2.2e+53
60	5.1e+116	4e+111	1.3e+104
80	5.1e+172	4e+167	1.3e+160
100	4.4e+232	3.4e+227	1.1e+220

Table 4.3: Number of total solutions with multiple seriation groups and processing time for sets of assemblages 4 < n < 100, testing solutions across 64 cores

feasibility of parallelized enumerative methods still explodes after 13 assemblages, but much more steeply. The goal of a new algorithm for deterministic multiple group seriations is, therefore, to employ heuristics to drastically reduce the size of the solution space. Vast amounts of the solution space involve partial orders which violate unimodality, but of course we cannot easily identify those regions of solution space *a priori* without testing possibilities. But given small partial solutions which do meet the seriation model, we can easily test solutions which are "adjacent" to the partial solutions, suggesting that agglomerative heuristics may be the best approach to finding a computationally feasible method.

CHAPTER 5

Measuring Cultural Relatedness Using Multiple Seriation Ordering Algorithms

Abstract Seriation is a long-standing archaeological method for relative dating that has proven effective in probing regional-scale patterns of inheritance, social networks, and cultural contact in their full spatiotemporal context. The orderings produced by seriation are produced by the continuity of class distributions and unimodality of class frequencies, properties that are related to social learning and transmission models studied by evolutionary archaeologists. Linking seriation to social learning and transmission enables one to consider ordering principles beyond the classic unimodal curve. Unimodality is a highly visible property that can be used to probe and measure the relationships between assemblages, and it was especially useful when seriation was accomplished with simple algorithms and manual effort. With modern algorithms and computing power, multiple ordering principles can be employed to better understand the spatiotemporal relations between assemblages. Ultimately, the expansion of seriation to additional ordering algorithms allows us an ability to more thoroughly explore underlying models of cultural contact, social networks, and modes of

social learning. In this paper, we review our progress to date in extending seriation to multiple ordering algorithms, with examples from Eastern North America and Oceania.

Source Submission to Electronic Symposium, "Evolutionary Archaeologies: New Approaches, Methods, And Empirical Sufficiency" at the Society for American Archaeology conference, April 2016 Co-authored with Carl P. Lipo.

5.1 Introduction

Seriation is a set of methods that uses patterns in the occurrence or abundance of historical classes to construct an ordering among otherwise unordered assemblages or objects (Dunnell, 1970). Its early 20th century developers built seriation as a relative dating method and orders constructed by seriation were intended to be chronological (O'Brien and Lyman, 2000, 1998; Lyman and O'Brien, 2006b; O'Brien and Lyman, 1999b; Lyman et al., 1997). While practitioners such as James Ford (Ford, 1938; Phillips et al., 1951; Ford, 1935) noted that seriation techniques also create orderings which incorporate the effects of spatial variation in addition to temporal change, the dominant use of seriation in archaeology has been to build chronology.

As a chronological tool, seriation has been success in developing an understanding the large-scale temporal structure of the archaeological record in the New World (Beals et al., 1945; Bluhm, 1951; Evans, 1955; Ford, 1949; Kidder, 1917; Mayer-Oakes, 1955; Meggers and Evans, 1957; Phillips et al., 1951; Rouse, 1939; Smith, 1950). Despite this success, the method has largely been ignored since the advent of radiocarbon dating given its primary association as a relative dating method. But seriation is only a dating method in the sense that chronology is one possible inference that can be obtained by mapping the spatiotemporal pattern of change in cultural variants. Other inferences are possible, and in particular, there is a growing understanding that seriation is one of several methods for documenting the evolutionary history of past populations (Lipo et al., 1997; Lipo and Madsen, 2001; Lipo, 2001b,a, 2005; Lipo and Madsen, 1997; Lipo et al., 2015; Neiman, 1995; O'Brien and Lyman, 1999b; Teltser, 1995).

Seriation is based on the notion that the frequencies of classes of artifacts reflect heritable continuity when it arises from information being passed between populations over time; that is, from cultural transmission processes. Although the fact that seriation, in some sense, measures cultural transmission has been implicit since the earliest discussions of the method (e.g., Kroeber, 1923), the connection remained a common sense generalization until the mid 1990's. Fraser Neiman (1990), in his dissertation and later his seminal article (Neiman, 1995), noted that the unimodal patterns that form the core of the traditional frequency seriation technique are regularly seen in the trajectories seen when simulating unbiased transmission. In order to make this connection both rigorous and useful in empirical work, we began a research program aimed at exploring the connection between cultural transmission models and seriation methods (Lipo et al., 1997). Our investigation into seriation has resulted in numerous publications, new seriation software algorithms, and many conference papers (Lipo and Eerkens, 2008; Lipo and Madsen, 2001; Lipo, 2001a, 2005; Lipo and Madsen, 1997; Lipo et al., 2015; Madsen and Lipo, 2014, 2015; Madsen et al., 2008; O'Brien et al., 2015).

The core of the all seriation techniques are a set of "ordering principles" which describe how the data points making up each assemblage or object are rearranged in order to achieve a valid seriation solution. Traditionally, there are two principles: occurrence and frequency (Dunnell, 1970; Rouse, 1967; Whitlam, 1981). The "occurrence principle" states that a valid ordering leaves no temporal gaps in the distribution of the historical classes used, and thus that temporal orders are continuous (Dempsey and Baumhoff, 1963; Rowe, 1959). The "frequency" or "popularity" principle states that in a valid ordering, the frequencies making up the continuous distribution of each historical type will be unimodal, possessing a single peak of "popularity" (Nelson, 1916).

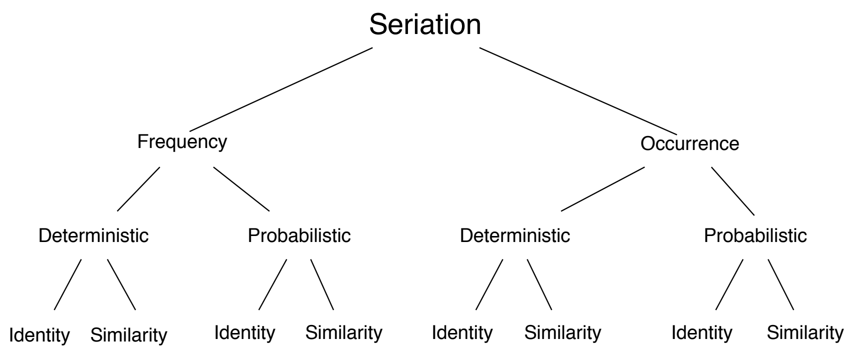


Figure 5.1: Dunnell (1981) defines seriation to be a set of methods which use historical classes to chronologically order otherwise unordered archaeological assemblages and/or objects. Historical classes are those which display more variability through time than through space. Occurrence seriation uses presence/absence data for each historical class from each assemblage (Kroeber, 1916; Petrie, 1899). Frequency seriation uses ratio level abundance information for historical classes (Spier, 1917; Ford, 1935, 1962). Frequency and occurrence seriation techniques can take the form of deterministic algorithms that require an exact match with the unimodal model or probabilistic algorithms that accept departures from an exact fit. Identity approaches employ raw data (whether frequency or occurrence) to perform the ordering. Similarity approaches transform the raw data into a non-unique coefficient (e.g., Brainerd Robinson, squared Euclidean distance); the coefficients then form the basis for ordering.

Both the frequency and occurrence principle work to sort descriptions of assemblages through time. The robustness of methods built on these principles is easily demonstrated by the continued utility of the basic chronological frameworks erected by culture historians in the first half of the 20th century using seriation along with stratigraphy and marker types (Lyman et al., 1997). It is intriguing to note, however, that the frequency principle remains an empirical generalization that is only suggested by the *generalized* behavior of cultural transmission models. Unimodality is not a necessary consequence. From Neiman's simulations (i.e., Neiman, 1995), one can see that the results of cultural transmission are not strictly or necessarily unimodal. This possibility suggests to us that seriation as a method requires further methodological development, especially if it is to be one of our major tools in tracing historical and heritable continuity in the archaeological record.[1]

In this paper, we explore an alternative to unimodality and the "popularity principle" that drives classical frequency seriation: exact minimization of inter-assemblage distance metrics, or "continuity" seriation. Although not a new principle, it was underappreciated especially prior to the contemporary explosion of computing power. We demonstrate that an exact form of distance minimization, in contrast to the statistical or approximate minimization associated with multidimensional scaling, generates solutions that are usually identical to the application of unimodality to the same data. Furthermore, using simulated data, we examine situations where frequency and continuity seriations may differ in minor ways, without affecting the overall ordering of the data set. Although there is still great value in the classical approach, the advantage of developing new seriation approaches is that we can often apply distance minimization to classes and types which do not necessarily display the classical unimodal form, which opens seriation to wider classes of data. In addition, distance minimization can be formulated within large scale, parallel machine learning frameworks, and thus made applicable to large data sets, increasing our ability to examine regional scale phenomena in archaeology, and making seriation useful for detailed analysis of contemporary data sets which dwarf the sample sizes typically available to archaeologists.

5.2 Seriation and the Frequency Principle

Seriation, in the Americanist sense, was initially developed by Alfred Kroeber (Kroeber, 1916) in the Southwest, based on his observations of changes in the relative abundance of forms of ceramic decorations found on sherds located in assemblages near Zuni Pueblo. The primitive seriation proposed by Kroeber was quickly amended by Leslie Spier, Alfred V. Kidder and Nels C. Nelson all of whom

---

[1] Cladistics and phylogenetic methods, especially those which take into account temporal differences in the samples being studied (stratocladistics) and which are capable of yielding phylogenetic networks in addition to trees, are the other major tools by which we can measure heritable and historical continuity.

were conducting stratigraphic excavations in the American Southwest (Kidder, 1917; Nelson, 1916; Spier, 1917). This group of researchers all noticed that when ceramics were described in a particular way – called "stylistic" by Kidder (1917) – the temporal distribution of the types took the form of "normal curves." Using such types, it was apparent that a series of assemblages collected from the surface or otherwise undated could be arranged in chronological order by rearranging them so that all type distributions approximated "normal curves" simultaneously. The orders constructed in this way could also be tested by finding stratified deposits and were found to be correct. The resulting method then went on to dominate archaeological practice for much of the next 50 years (Lyman et al., 1997).

As powerful as seriation proved to be, these early formulations were entirely intuitive and based on the generalization that greater temporal differences between assemblages caused larger differences between frequencies of decorated types, and that properly constructed historical types displayed a clear pattern of change (Phillips et al., 1951, p. 220):

> If our pottery types are successful measuring units for a continuous stream of changing cultural ideas, it follows that when the relative popularity of these types is graphed through time, a more or less long, single-peak curve will usually result. Put in another way, a type will first appear in very small percentages, will gradually increase to its maximum popularity, and then, as it is replaced by its succeeding type, will gradually decrease and disappear.

This compactly describes the "popularity principle," originally articulated by Nelson (1916) and Wissler (1916). A key word in the above is "usually," since not all types display the unimodal distribution described, even when the attributes chosen are explicitly stylistic and decorative. Types suitable for frequency seriation were a subset of stylistic variation, comprising those which displayed spatial and temporal contiguity, a long enough duration that the types overlapped in their representation among sites and assemblages, and those whose distribution through time displayed the characteristic unimodal form which allowed the analyst to arrange them by eye. Culture historians also minimized the effect of chance and potential recurrence by insisting that the classes used for measurement were constructed from multiple dimensions (Phillips et al., 1951; Lipo, 2001b). The overall process of constructing and testing such types became known, after Krieger (1944), as applying the "test of historical significance."

5.2.1 Unimodality and Cultural Transmission Processes

In most cases (such as the above quote from Phillips, Ford, and Griffin), the popularity principle is simply assumed to hold in culture-historical applications. It is clear that culture historians assumed that what generates heritable continuity, and thus allows the tracing of chronological relations, is cultural transmission. As Lyman (2008) documents in careful detail, early 20th century anthropology and archaeology understood and discussed a variety of transmission processes informally, as generating the patterns they studied, even if they used different terms and did not form quantitative models for it. Rouse (1939), for example, explicitly discussed the diffusion of cultural traits, in terms that we now recognize as a spatiotemporal model of transmission. Kroeber, the father of frequency seriation, clearly understood the connection between his previous work and trait diffusion (Kroeber, 1937). Deetz and Dethlefsen (1965; 1971) noted the spatial dimension to trait diffusion. There are many more examples (Lyman, 2008).

Interest in studying cultural transmission in an explicit way has a long history in archaeology. Since the 1970s, archaeologists have worked with models of diffusion, with those models becoming increasingly quantitative, statistical, and even explicitly mathematical (e.g., Ammerman and Cavalli-Sforza, 1971). These models of diffusion, however, tended to be deterministic, especially those stemming from the interdisciplinary literature on the diffusion of innovations (e.g., Rogers, 2003). Deterministic models, however, ignore the essential historically contingent pathways of culture transmission that produce the patterns noted by culture historians as historically significant. More recently archaeologists have become interested in developing models for individual social learning events (see discussion in Chapter 3.

It was not until archaeologists began working with stochastic models of cultural transmission, however, that we could easily visualize the sheer variety of patterns that cultural transmission processes can, and do, generate. Stochastic models of transmission allow us to easily explore the precise conditions under which unimodal distributions occur in type frequencies, what classification methods tend to produce it, and what dimensions of variation combine to produce mostly unimodal behavior.

Dunnell's (1978) exposition of style as neutral variation was one key step in the adoption of stochastic models of drift from population genetics as the main tool for exploring cultural transmission dynamics. Neiman (1995) took this step substantially further when he simulated drift in cultural variants as an unbiased transmission process, as shown in Figure 5.2. Immediately apparent is the fact that some variants do display unimodal patterns, but most variants are multimodal or display violations of unimodality at small scales even if the macroscopic shape seems to conform to the popularity principle.

Figure 5.2 shows that there is nothing necessary about unimodality given cultural transmission. Instead, culture-historical classifications and typologies were constructed such that they produced compact spatiotemporal distributions and generally followed unimodal histories. This is precisely what Krieger's (1944) "test of historical significance" yields when applied to a candidate typology. This is accomplished by ensuring that types are composed of multiple dimensions of variation which co-occur on artifacts identified to that type. Each dimension of variation (e.g., surface treatment)

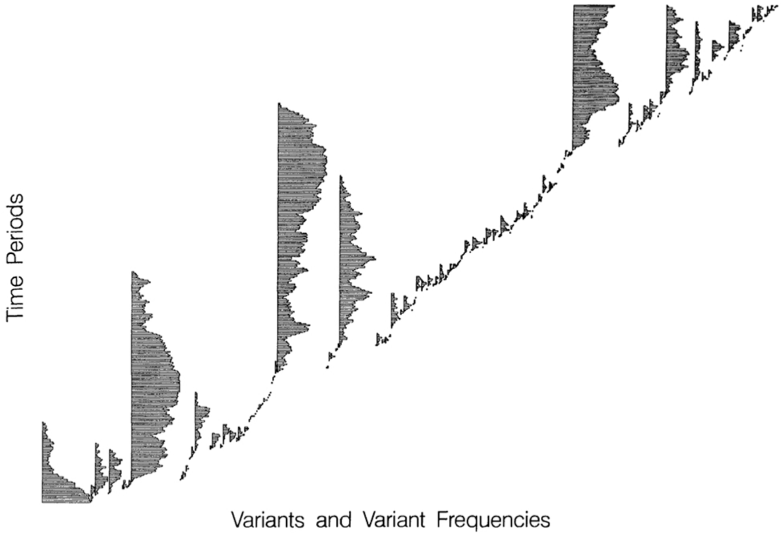


Figure 5.2: Neiman's simulation of drift in cultural variant frequencies under unbiased cultural transmission (reproduction of Figure 2a from Neiman 1995.)

may have complex histories, like those seen in Figure 5.2, but when we combine several dimensions into a class, the history of the *co-occurrence* of each combination of attributes becomes smoother and more localized in time and space. This process of class construction necessarily results in a more compact spatiotemporal distribution for the class than for any of its constituent attributes.

Unimodality further arises from the constructed classes by examining their *relative* frequency over time, expressed in percentages or fractions, which sum to 100% or 1.0. The latter constraint, which causes some combinations to be relatively less numerous when others increase in prevalence, smooths the more complex variation of individual attributes into the histories of "popularity" that early twentieth century anthropologists noted (e.g., Nelson, 1916; Wissler, 1916).

Taken together, these factors seem to explain why the intuitive construction of historical types, from the continuous flow of the products of cultural transmission processes, worked to produce chronology through application of the common-sense popularity principle, and why not all artifact classes constructed from otherwise "stylistic" dimensions of variation, are suitable for frequency seriation using unimodality as the ordering criterion. From the perspective of culture historians, unimodality was a sufficient criteria for recognizing patterns that were likely chronological from those that were likely not. While focusing on only those classes that produced unimodal distributions in class frequencies might have ignored other potentially historical significant classes, without any other means of identifying chronological patterns, culture historians were satisfied with the subset that worked.

5.2.2 Continuity: An Alternative to Unimodality

There are several reasons why we should explore alternatives to unimodality as an ordering algorithm for frequency seriation. First, from a performance perspective, searching for unimodal orders is computationally expensive, even for relatively small data sets (Madsen and Lipo, 2014). Even with the iterative, agglomerative method that we introduced recently (Lipo et al., 2015), the computation time can grossly exceed computing capacity for data sets as small as 30. While 30 is a large number of assemblages by most archaeological standards especially when adequate sample size requirements are met, it is a serious limitation. Without good techniques and ordering principles seriation may not scale to much larger problems, and even be applicable to the flood of data seen in modern day life.

Second, and more importantly from a theoretical perspective, it is important to be able to trace heritable continuity even if does not display a particular type of temporal frequency distribution. Using traditional type construction methods and the test of historical significance, culture historians were able to find *enough* conforming types and classes to construct regional chronologies. The goal of culture historians was to build chronologies using the most efficient means possible to do so, not study combinations of trait transmission through time and space. The use of seriation as a method for tracing evolutionary relationships is a more demanding task than establishing rough chronology in a region. Thus, it is worth searching for additional ordering principles that may be useful for seriating more classes of cultural variants. Specifically, there is strong relationship between the number of classes in a seriation, and our ability to map differences across space and time. We need methods that can evaluate arbitrary sets of classes to arrive at the most detailed understanding of cultural transmission landscapes (see Chapter 6).

A theoretically sound ordering principle for seriation should be derivable from characteristics of the underlying cultural transmission processes that we believe drive the spatiotemporal variation seriation measures. Formal models of cultural transmission, such as those formulated by Boyd and Richerson, Cavalli-Sforza and Feldman, and borrowed from population genetics (Boyd and Richerson, 1985; Cavalli-Sforza and Feldman, 1981; Neiman, 1995) provide a good starting place. Their models incorporate stochastic autoregressive processes in which the probability distribution of outcomes at a given time are dependent upon the outcomes from the immediate past. Mathematically, then we can treat cultural transmission models as Markov processes, usually of first order (i.e., without dependencies on states previous to the immediate past state). Such models are certainly capable of making large changes in state over short time intervals, but large jumps are rare compared to small changes in state, especially in large populations. This is the reason why we (and culture historians) often have an expectation that cultural transmission has a "gradual" character to it.

The probabilistic gradualism of change over small time periods in our cultural transmission processes explains the "continuity" principle that is embedded in traditional forms of seriation. Continuity is strongly related to notions of continuous functions in mathematics: samples which originate close together in time, space, or both will be close in type frequency and the presence/absence of types, especially compared to samples which are further apart. This continuity principle immediately leads to considering ordering algorithms based upon minimizing a suitable distance metric, with assemblages represented by points in a multidimensional space of type frequencies or counts.

5.2.3 Statistical Seriation Methods

The earliest statistical techniques for seriation were also built upon using interassemblage distance metrics. Brainerd (1951) and Robinson (1951) pioneered a method for seriation based upon the similarity between assemblages, measured as a scaled version of the Manhattan (or city-block) distance between assemblage frequencies. When these scaled distances (which became known as Brainerd-Robinson coefficients) are arranged in a matrix with the largest values nearest the diagonal and the lowest values in the corners and away from the diagonal, the order of assemblages by row or column provides the seriation solution. In practice, most real data matrices cannot be put in perfect Robinson form without violations from the assumptions of the seriation model.

Brainerd and Robinson's pioneering work became the basis of large literature devoted to creating methods for matrix ordering in the face of the practical difficulties in coercing most data sets into a linear ordering (e.g., Dempsey and Baumhoff, 1963; Kendall, 1963; Matthews, 1963; Bordaz and Bordaz, 1970; Gardin, 1970; Kendall, 1970, 1971). As access to computers by researchers in the social sciences increased, computerized algorithms for examining permutations quickly proliferated (Ascher and Ascher, 1963; Craytor and Johnson, 1968; Kuzara et al., 1966). Kendall (1969) and others attacked the ordering problem through the use of multidimensional scaling. For a detailed review of the many variants on this type of probabilistic seriation solution through the late 1970s, see (Marquardt, 1978). Most recently correspondence analysis has been used with success in determining probabilistic seriation orders, and even more importantly, quantifying the degree of departure from the ideal seriation model (Smith and Neiman, 2005).

There have been calls to simplify the problem by directly minimizing inter-assemblage distance, and thus the total "path length" of a candidate seriation solution. Kadane (1971) describes this approach, and it was adopted later by Shepardson (2006) in his construction of the "Optipath" seriation algorithm, which has distance minimization at its core. We agree with Kadane and Shepardson and explore the concept in the next section.

Where existing distance/similarity methods encounter a problem is the assumption that a seriation solution must be a single linear order. This assumption was made for two reasons. First, seriations were seen as offering relative chronology, which a linear order idealized for the analyst. Second, even though practitioners understood that there was spatial variation in the histories of class frequencies, there was no feasible manual method for finding more complex multidimensional solutions. The advent of digital and then personal computers solved the latter problem, opening up the possibility of using *all* of the information at our disposal about space and time to discover the multiple historical relationships present in our data.

In an earlier paper, we describe a seriation algorithm (iterative deterministic seriation solutions, or IDSS) that finds all of the possible orders in a set of data that conform to an ordering principle, and where those orders have overlap in assemblages (Lipo et al., 2015). Using this ordering principle, IDSS constructs a graph with branches that recognizes that the best solutions may not be linear. In probabilistic approaches to seriation such as MDS or correspondence analysis, departures from linear solutions have always been treated as "stress" or "error." Practitioners usually recognize that such departures arise from coercing data which naturally sit in a larger number of dimensions – because of spatial variation and other factors – into a one-dimensional order. In essence, methods which attempt to coerce a complex spatiotemporal pattern into a linear ordering tend to treat departures from linearity as noise, which is then ignored.

But the departure from linearity is not "noise," in the statistical sense. Especially if one accounts for sampling error in constructing seriation orders (as we do in IDSS by using the bootstrap to construct confidence intervals around the empirical frequencies), then departures from a linear ordering are *signal*, not noise. Such solutions reflect the fact that an assemblage at time T_1, for example, may be the closest match to two different assemblages at later times T_2 and T_3 for example, given slightly different areas of overlap in their type frequencies. This pattern can occur because the seriation method is inherently spatiotemporal, instead of simply measuring time (as culture historians have always known), and it can also reflect the splitting of populations into separate lineages or their merger (see Chapter 6).

5.2.4 Exact Distance Minimization Ordering: "Continuity" Seriation

Instead of the "approximate" distance minimization algorithms employed in multidimensional scaling, we explore exact solutions using our IDSS algorithm (Lipo et al., 2015). For simplicity in the configuration of the software, we summarize our approach by calling it "continuity" seriation, to distinguish it from unimodal-based frequency seriation and to emphasize that we want solutions that have the smoothest, most continuous transition of type frequencies when we consider pairs of assemblages. We achieve this by locally minimizing the inter-assemblage distance within the solution graph, which automatically yields the minimum total "path length" for a seriation solution.

Our algorithm makes no use of the unimodality criterion, and produces equivalent results in almost all cases, as we show in the next section. The algorithm currently employs the Euclidean distance between assemblage counts or frequencies, although it can use any distance metric.[2] The Euclidean distance has the advantage of treating each class as equivalent measures, a property consistent with the use of paradigmatic classification (sensu Dunnell, 1971) for generating measurement classes. Given a table of inter-assemblage distance metrics, we first construct pairs of two-vertex graphs which represent the "closest" assemblage for each assemblage in the data set (mirrored pairs are filtered out since they are isomorphic). The edge weight given to each edge is the Euclidean distance between the assemblages represented by vertices. For each of the minimal graphs in this initial set, we then find the assemblage with the shortest distance to each of the two ends, and continue iter-

---

[2] We might, for example, want to explore ultrametric distances for classifications which incorporate dependency structure between traits, as described in Chapter 7 and Mesoudi and O'Brien (2008b)).

ating. Crucially, if there are equal-distance options, both possible solutions are retained. The result of this iteration is a collection of graphs which represent partial minimum-distance paths through the set of assemblages. This collection of partial graphs are then overlaid to form a single solution using a "minmax" approach as described in our paper on the IDSS algorithm in general (Lipo et al., 2015).

The general approach is the same one we take to frequency seriation in our our 2015 paper; what differs here with "continuity" seriation is how we form the set of candidate partial solutions. Instead of enforcing unimodality within each partial solution, we minimize Euclidean inter-assemblage distance. The resulting minmax graph is linear only if all of the candidate partial solutions perfectly overlay themselves into a linear solution, and otherwise will have a tree structure with branches. The possibility of branching is what allows a seriation solution to express both spatial and temporal structure simultaneously. The ability to inform on both allows investigation of social network structure, and interaction and social learning patterns in past populations, at scales more detailed than entire cultural manifestations or phases. We believe that seriation, augmented in this way, sits between the microevolutionary level where we investigate evolution in single populations, and the macroevolutionary level, best explored using the tools of phylogenetic analysis and cladistic techniques.

5.3 Comparing Frequency and Continuity Seriation

In this section we compare the results of our IDSS frequency seriation algorithm, described in a recent paper (Lipo et al., 2015), and seriations performed using exact distance-minimization or "continuity" algorithms. Ideally, we would wish for the faster distance-minimization algorithm to give the same results as employing the slower, more expensive algorithm employing unimodality as the ordering principle.

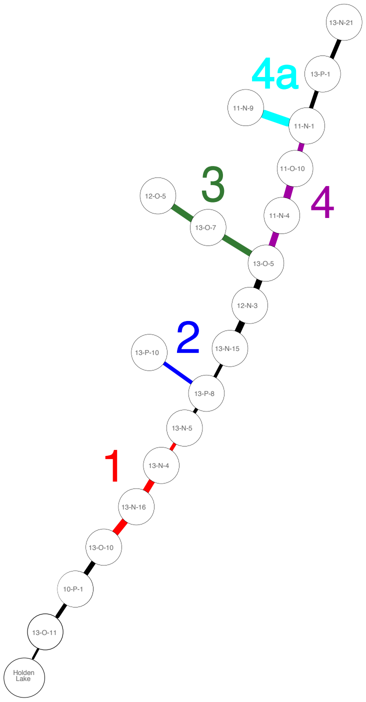


Figure 5.3: Seriation solution with frequency and continuity seriation for PFG (1951) ceramic assemblages in the Lower Mississippi River Valley, as analyzed by Lipo (2001a) and re-analyzed by Lipo et al. (2015). There are no differences between frequency and continuity ordering algorithms in analyzing this set of assemblages, and thus only one graph is shown.

To that end, we extended the Lower Mississippi River Valley example from our recent work (Lipo et al., 2015) by comparing frequency and continuity seriation algorithms on the same set of assemblages. The result is depicted in Figure 5.3. The result is identical – the two solutions are isomorphic.

We are archiving seriation data sets, with supporting information including geography if available, and licenses (if available and required) in an open-source repository on Github: `https://github.com/mmadsen/seriation-datasets`. Over time we expect to be able to employ the data sets to create automated test suites for new seriation algorithms and software implementations. We are also adding scripts which execute our IDSS software for any of its implemented algorithms, for the data sets. We welcome contributions to the data repository (please contact Madsen for more details).

Empirical data sets suitable for determining comparable the two seriation algorithms might be are hard to come by, and typical cases have a small number of assemblages, which tend to yield small linear seriations without complex structure. Thus, we also employ a Monte Carlo or simulation approach to sampling many possible seriations, performing both frequency and continuity on sampled data, and comparing the structure of the resulting seriations. We employed the *SeriationCT* simulation framework described in detail in Chapter 6 to generate samples of simulated cultural transmission, with time averaging, on realistic metapopulation-style models of regional interaction. The simulations were performed for purposes of examining whether seriations can be diagnostic of the evolving social networks within which cultural transmission occurred in the past. For purposes of the present study, however, the simulation results presented an opportunity to compare ordering algorithms.

Because the "lineage splitting" transmission scenario (described in Chapter 6, in Section 6.3.1) yielded seriation graphs with detailed branching structures, it presented a more conservative comparison than the complete networks, which had a much higher incidence of straight linear orderings. Thus, we randomly sampled 50 simulated data sets from the lineage splitting model, and performed seriations with our IDSS software, using both frequency (unimodality) and continuity (distance minimization) settings.

Seriation of artifact assemblages is inherently a regional-scale problem, whether for chronology or tracking interaction and social learning processes. Thus, the fundamental abstraction for modeling is a graph or network which (a) represents the intensity of contact, migration, and interaction between communities of people at any given point in time, (b) allows the set of communities to evolve, with some communities going away and others originating over time, and (c) representing how both the pattern and intensity of inter-community contacts evolves over time. Social network or graph models, especially weighted graphs, form an essential ingredient for this type of modeling, but need to be extended to the temporal dimension.

We measure whether frequency and continuity solutions are identical by testing whether the solution graphs are isomorphic, which means that the same vertices are connected to the same neighbors by the same edges. Of the 50 simulation runs examined here, in 80% of cases the continuity and frequency seriations give an exactly identical solution. Of the remaining non-identical solutions, we find that the differences nearly always involve the repositioning of a single assemblage. In the next section, we examine such a case in detail to understand what drives such differences when they occur.

5.3.1 Examining a Solution Which Differs

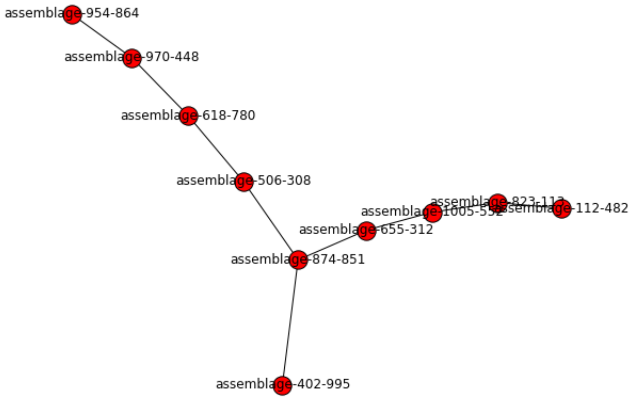


Figure 5.4: Frequency seriation solution for simulation run f8a6f378 on the ”lineage splitting” regional interaction model.

Of the differing solutions, we selected one (f8a6f378) at random to show the details of how frequency and continuity solutions differ. Figures 5.4 and 5.5 depict the frequency and continuity seriations, respectively, in the form of graphs which connect assemblages which are “adjacent” in the seriation solution. This makes it easier to see where an assemblage is really part of several solutions, which can indicate lineage splitting or differentiation occurring over space. We introduced this format for seriation solutions in our recent article on IDSS seriation (Lipo et al., 2015).

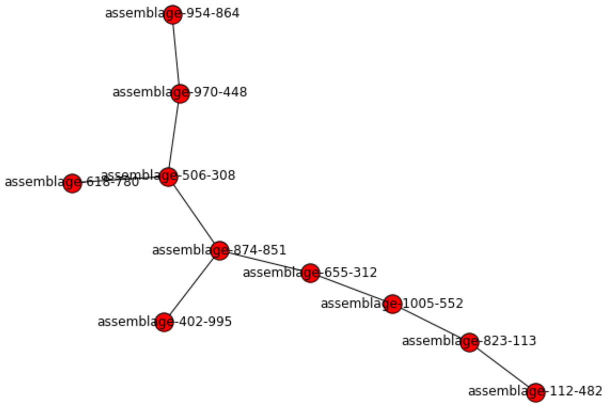


Figure 5.5: Continuity seriation solution for simulation run f8a6f378 on the ”lineage splitting” regional interaction model.

Although the graphs are laid out slightly differently (as a function of an automated graph layout algorithm), it is apparent that most of the seriation ordering is the same. Simulated assemblage 954-864 anchors one end of the ordering, while assemblage 112-482 anchors the other.[3] Both solutions also show a branch for assemblage 402-995, which belongs to one of the two lineages after the connections between two sets of communities is lost. It is a single assemblage branch because of the vagaries of sampling assemblages out of the total set of communities in this example.

The main difference between the solutions comes in assemblage 618-780 and where it connects. In the frequency solution it occurs “inline” while in the continuity solution, interassemblage distance is minimized by removing it to a small branch of its own.

Viewed in traditional tabular view of the type counts in Tables 5.1 and 5.2 or as traditional centered bar charts in Figures 5.6 and 5.7, the reason for this difference in assignment for assemblage

---

[3] Simulated assemblage names here reflect geographic coordinates, since regional interaction models often bias interaction and migration by location or neighborhood.

Assemblage Name	6022-0-1767	36526	36557	7005-0-1767	7628-0-1767	0-9222-3	1-0-1767	3771	6996-4-3
assemblage-954-864	10	160	0	49	92	0	0	0	9
assemblage-970-448	0	155	0	74	128	0	0	0	14
assemblage-618-780	123	50	0	164	121	0	13	0	14
assemblage-506-308	107	58	0	199	114	0	9	0	13
assemblage-874-851	81	66	0	165	0	0	162	6	17
assemblage-874-851	81	66	0	165	0	0	162	6	17
assemblage-655-312	0	52	16	111	0	20	269	6	26
assemblage-1005-552	0	53	32	72	0	61	182	41	8
assemblage-823-113	0	145	81	0	0	64	132	10	14
assemblage-112-482	0	24	151	0	0	157	81	49	9
assemblage-874-851	81	66	0	165	0	0	162	6	17
assemblage-402-995	106	65	0	29	0	0	192	0	7

Table 5.1: Raw data for frequency seriation for simulation run f8a6f378, grouped into blocks corresponding to the branches of the solution graph

618-780 can be seen. First, there are apparent violations of unimodality in the frequency seriation. If these violations exceeded the calculated confidence limits around each frequency, then 618-780 would have been pulled out of the ordering and placed into a separate solution branching from assemblage 506-308. But in this case, the frequencies overlapped in their confidence intervals and there was no reason to pull it out and form a separate solution branch.

In the continuity seriation result, which does not employ confidence intervals around frequencies at present, the interassemblage distances in frequencies clearly pick up on the same mismatch. Our initial continuity algorithm then moves assemblage 618-780 into a separate solution branch. Clearly, the sampling error that motivated use of confidence intervals in the "unimodal frequency" seriation case apply here as well, but in our initial algorithm work it was not obvious how to incorporate intervals into the interassemblage distance calculations. That remains an open question, but an important one for demonstrating the full equivalence of methods.

5.4 Discussion

The fact that distance minimization can function as a seriation ordering algorithm is not a new idea. Not only has there been development of the idea within archaeological circles in the work of Kadane, Shepherdson, and others, but distance minimization of one type or another underpins most classical multivariate statistics and nearly all of contemporary machine learning. Our principal contributions here have been to explicate the relationship between different seriation ordering algorithms, and to reintroduce distance minimization in an "exact" rather than statistical form.

Exact distance minimization as a means of tracing patterns of cultural transmission is only possible if we do not coerce the data into a single linear ordering, as has been the practice in all previous work. In these previous applications, the departures from linearity have been considered statistical

Assemblage Name	6022-0-1767	36526	36557	7005-0-1767	7628-0-1767	0-9222-3	1-0-1767	3771	6996-4-3
assemblage-954-864	10	160	0	49	92	0	0	0	9
assemblage-970-448	0	155	0	74	128	0	0	0	14
assemblage-506-308	107	58	0	199	114	0	9	0	13
assemblage-874-851	81	66	0	165	0	0	162	6	17
assemblage-655-312	0	52	16	111	0	20	269	6	26
assemblage-1005-552	0	53	32	72	0	61	182	41	8
assemblage-823-113	0	145	81	0	0	64	132	10	14
assemblage-112-482	0	24	151	0	0	157	81	49	9
assemblage-874-851	81	66	0	165	0	0	162	6	17
assemblage-402-995	106	65	0	29	0	0	192	0	7
assemblage-506-308	107	58	0	199	114	0	9	0	13
assemblage-618-780	123	50	0	164	121	0	13	0	14

Table 5.2: Raw data for continuity seriation for simulation run f8a6f378, grouped into blocks corresponding to the branches of the solution graph

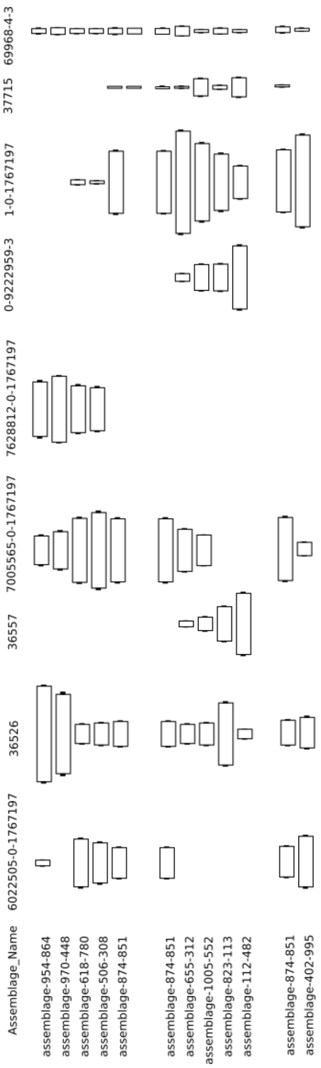


Figure 5.6: Centered bar chart representation of the relative frequencies of type for simulation run f8a6f378 built with the IDSS frequency seriation algorithm. The groups correspond to the branches of the solution graph.

noise or "stress," and disregarded. From a culture transmission model, however, noise only enters the seriation problem as sampling error of counts or frequencies given the size of sample taken by the analyst. We can control this type of noise by using bootstrap confidence intervals around the empirical frequencies when we make ordering decisions. Our IDSS software system does so by default.

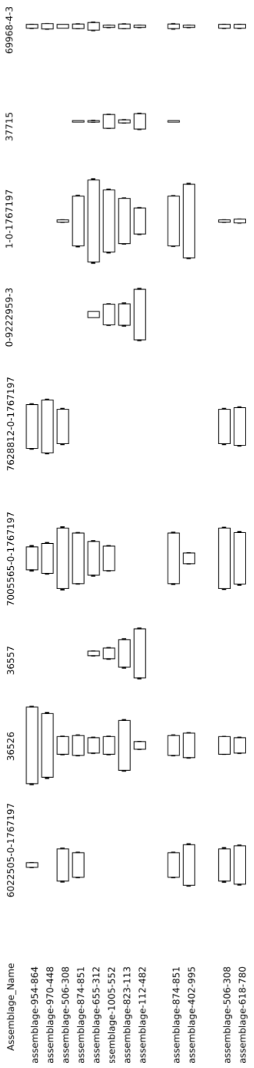


Figure 5.7: Centered bar chart representation of the relative frequencies of type for simulation run f8a6f378 built with the IDSS continuity seriation algorithm. The groups correspond to the branches of the solution graph.

Thus, once the effects of sampling are controlled departures from linearity cannot be noise, but are telling us something else about our data. In our judgment, those departures from perfect linearity are telling us about the simultaneous effects of spatial variation, temporal order, and the structure of the social networks of interaction within which past cultural transmission occurred.

Thus, our approach to both frequency and continuity seriation allows partial solutions (each of which **is** a valid linear ordering) to agglomerate to form graphs or networks of solutions, given vertices (assemblages) which overlap between the sub-solutions. The resulting seriation graphs give us a more complete picture of the multiple causes that drive seriations than do traditional linear orders, whether perfect or coerced by a statistical method.

The search for additional ordering methods led us to reconsider distance minimization methods, and although it is not unexpected that such methods work, it is a happy result. Continuity techniques have a much lower computational burden than searching for unimodality, especially as the number of assemblages gets large. For the Phillips, Ford and Griffin assemblages discussed here, the frequency solution took 25.2 seconds on an 8 core system, while continuity analysis took 0.955 seconds, for a speedup of 26x. This performance difference should be taken as a minimum on the difference between algorithms, because our current algorithm for unimodality analysis is parallelized for a critical section across all of those cores, while continuity is still a serial algorithm and only uses a single core. Realistically, we should see a much larger speedup with further development, especially given the wealth of parallel algorithms for distance metric computations in contemporary machine learning. The latter will allow continuity methods to be fruitfully used even for "big" datasets of the type easily gathered in online settings. This method effectively has no limit as to the number of assemblages that can be analyzed.

Seriation is among the oldest of the purely archaeological methods for determining both chronology and cultural relatedness, but we find that it continues to repay detailed exploration by archaeologists and students of cultural evolution. It is fully complementary to phylogenetic methods and cladistics in many ways, especially in its ability to use detailed information about trait abundances and the spatial pattern of those abundances instead of largely presence/absence data on character states. This makes seriation, in our view, the method of choice for "mesoscale" problems and questions.

CHAPTER 6

A Computational Method for Identifying Regional Interaction Patterns From Seriation Solutions

Source Paper delivered at the Human Behavior and Evolution Society Annual Meeting, 2016, in a session ”Macroevolutionary Approaches to Cultural and Technological Evolution.” [1]

6.1 Introduction

One of the most important tasks of an evolutionary archaeology is documenting evolutionary history itself. This means more than creating chronology and documenting the history of material culture change, although those are essential tasks. Without the “chronicle” (O’Hara, 1988) of empirical facts, we cannot construct the testable narratives that represent the evolution of culture. But evolutionary history requires more than documenting chronology and change in artifact types and assemblages; it requires inference of *heritable continuity* from the “mere facts” of *historical continuity*

---

[1] Archived as `https://figshare.com/articles/madsen2016-hbes-computational-interaction-patterns-slides_pdf/3468650`.

in artifact classes (O'Brien and Lyman, 2000). Doing so requires separating similarity which represents *homology*—similarity due to descent—from convergence due to natural selection but without common descent (Dunnell, 1978; Kroeber, 1931).

Dunnell (1978; 1980; 1989) argued that despite lacking a real theoretical basis for the endeavor, early culture historians had developed intuitive methods for tracing heritable continuity (O'Brien and Lyman, 2000; Lyman et al., 1997). In modern terms, the core methods of culture history such as Krieger's (1944) "test of historical significance" and seriation (Dunnell, 1970; Ford, 1935, 1936, 1949) allow the tracing not just of chronology, but provides evidence of heritable continuity and thus the history of past cultural transmission (Lyman, 2008).

At the largest spatiotemporal scales, phylogenetic methods are replacing the traditional tools of culture history for framing hypotheses about the chronicle of cultural change at the largest scales (Borgerhoff Mulder et al., 2006; Lyman et al., 1997; Lyman and O'Brien, 2006a; O'Brien and Lyman, 1999a; O'Brien et al., 2000, 2001, 2003; O'Brien and Lyman, 2003; O'Brian and Lyman, 2000; Prentiss and Laue, 2019; Prentiss et al., 2015; Tëmkin and Eldredge, 2007). The "resolution" of methods like phylogenetic analysis depends upon the nature of the classifications employed, and the fact that analyses typical employ only presence/absence data rather than detailed class frequencies. Especially because class presence/absence is the typical criterion for splitting on trees, most uses of phylogenetic methods in archaeology are "macroevolutionary" (Prentiss and Laue, 2019).

When we wish to operate at smaller scales and over shorter amounts of time, we need methods that employ not just presence/absence information about classes and types, but all of the frequency and spatial information at our disposal.[2] This is especially true in the archaeology of the late Holocene, when we wish to resolve events that might represent change over decades rather than centuries, and within smaller study areas. This level of detail has long been part of archaeology, with

---

[2] This is not to say that cladistic methods are not applicable to mesoscale analysis, but they may require custom classifications with finer levels of "splitting" of attributes.

regional studies and syntheses like Ford's work in the Virú Valley of Peru (Ford, 1949) and Phillips, Ford, and Griffin's (1951) study of the Lower Mississippi River Valley. More recently under the "New Archaeology," work at this scale was often included under the general umbrella of "regional analysis" (Johnson, 1977). When focused on tracing the history of cultural transmission, at the spatial and temporal scales mentioned, I call this kind of analysis "mesoscopic," since it lives in an intermediate zone between the "microevolutionary" study of single assemblages or localities, and the "macroevolutionary" focus of much larger scale and comparative work.

The core methods of culture history, and especially seriation, were designed precisely to answer questions about chronology and cultural transmission at mesoscopic and macroevolutionary scales, but in intuitive ways with common-sensical explanatory concepts such as "trade" and "migration" (O'Brien and Lyman, 2000; Lyman et al., 1997; Lyman, 2008). Because the end products were atheoretical and interpretive, there was little reason to develop a more detailed quantitative understanding of what could be done with methods like seriation, and when radiometric dating methods arrived on the scene, it was easy to frame seriation as nothing more than a "relative dating method" instead of a generalized method for tracing homology, *which just happened* to deliver temporal information as a side effect.

This paper builds upon previous work by myself and Carl Lipo reformulating seriation in a rigorous way to focus on the tracing of homology and construction of evolutionary chronicles (Lipo et al., 1997; Lipo, 2001b; Lipo et al., 2015; Madsen et al., 2008) and Chapter 5. In this work, I focus on the question of how seriation graphs, output from our iterative deterministic seriation algorithm (IDSS), can function as observable data in fitting scenarios for the history of cultural transmission at regional scales. The goal is to build a computational and statistical approach to describe: (*a*) the structure of cultural relatedness within a region, and (*b*) our hypotheses about the cultural transmission history in quantitative ways that allow us to use standard machine learning and statistical tools to perform model selection and assess goodness-of-fit.

This paper introduces "interval temporal networks" as the natural data structure for representing our hypotheses about the history of cultural transmission. Simulation of cultural transmission on interval temporal networks provides samples of the distribution of cultural traits we would expect to see for a given hypothesis. Seriation graphs can be constructed both for the simulation output of each hypothesis, and for our empirical data from archaeological assemblages. We then extract summary graph metrics for seriation graphs, allowing us to describe their structure in quantitative terms suitable for a machine learning classifier to be used to assess: (*a*) the identifiability of different cultural transmission hypotheses given seriation graphs (i.e., their equifinality between hypotheses), and (*b*) the transmission hypothesis for which our empirical seriation graphs have the closest match.

The results reported here are preliminary but encouraging. Using seriations as the observable variable alone, it is possible to differentiate between regional transmission histories where communities are largely balanced in their interaction and interact with all their peers, from histories that show very localized, "nearest neighbor" interaction, from histories involving the coalescence of different lineages, or their splitting to form two or more groups which no longer interact. I examine these scenarios with respect to our previous seriation results for Mississippian ceramic assemblages in the Central and Lower Mississippi River Valley.

6.2 Documenting the Regional History of Cultural Transmission With Seriation Graphs

The culture-historical seriation method employs an ordering criterion, either continuity of class presence/absence, or unimodality of class frequencies, to provide a relative chronological order. It does so under an assumption that the "correct" order will be the one that introduces no discontinuities in any of the classes used in the ordering (Ford, 1949; Phillips et al., 1951; Rouse, 1939). This means that in occurrence seriation, the correct order will not contain gaps where classes exist in a region, go away, and then come back. Instead, all classes will display continuous existence throughout their respective lifetimes. For frequency seriation, the correct order will not display situations where some classes have discontinuous "jumps" in frequency value inconsistent with their local trend. These assumptions are what one would expect if seriation is providing a relative history of how different cultural traits, expressed as archaeological classes or types, were learned and inherited within a population or populations in a region.

Several things can disturb our ability to cleanly order a set of samples into a single linear, chronological relationship. Dunnell (1970) describes these, and the criteria traditionally used to minimize them and produce usable seriations. First, we expect all of the assemblages to represent comparable durations of time, otherwise some will be time averaged over a longer history of transmission and cultural change than others, and not be orderable with the rest. Second, in a traditional seriation we minimize the amount of spatial variation involved, since cultural traits flow across space as well as time. If we want seriations to only depict the chronological relationships, we must minimize the amount of spatial variation we include (Phillips et al., 1951; Rouse, 1967).

If our goal, on the other hand, is not to extract just chronology, but understand the spatiotemporal history of how cultural traits spread within a region, then we need to do several things. First, we need to stop trying to force the seriation to yield a single linear order. Instead, if there are multiple valid ways that the frequencies of the classes can be arranged, we should express these as individual possible "solutions". In our 1997 paper with Tim Hunt and Robert Dunnell (1997), we proposed breaking seriation solutions into multiple subsets, with the unimodality (or occurrence) criterion applied strictly within each subset. This allows deterministic solutions (*sensu* Dunnell 1970, 1981) which are temporal within each subset, while the differences between subsets reflects spatial variation in the history of cultural traits. Some classes may increase in frequency spreading to the north, for example, at the same time that the same classes are declining in frequency in a southern direction. Attempting to incorporate both patterns in the same linear ordering will result in no solution (or, in the case of multivariate statistical methods, a low confidence solution with much "noise" whose utility is entirely unclear).

Lipo (2001b) took another step towards formalizing seriation in his dissertation by investigating methods for assessing the quality of a proposed solution. He employed a bootstrap resampling approach to estimate the likely standard error around frequency estimates given the overall sample size we possess for a site or assemblage. This allows us to assign a confidence interval to each "bar" in a frequency seriation. Confidence intervals can then be used to: (*a*) perform pairwise comparisons between assemblages to assess the significance of an ordering, and (*b*) understand quantitatively which assemblages are "contemporaneous" in the sense that the data at hand cannot distinguish between them.

The ability to test significance of ordering decisions and detect contemporaneity provided the tools to perform an automated breakdown of seriation solutions into multiple solutions, as we proposed in our early work to capture both spatial and temporal information. In our 2015 paper describing the IDSS algorithm (Lipo et al., 2015), we took the additional step of representing the relationship between each subset in a seriation with multiple possible solutions. Whenever an assemblage is present in two or more subset solutions, we link the linear subsets together into a graph, at the assemblage which appears in both. Figure 6.1 shows an example of such a graph, which merges two possible solutions from a simulation of cultural transmission with spatial structure.

The ability to use seriations as observable data, to which we directly fit hypotheses about the history of cultural transmission, requires that seriations be large enough, and encompass enough spatial and temporal variation, that there is meaningful *structure* to the resulting graphs. If every

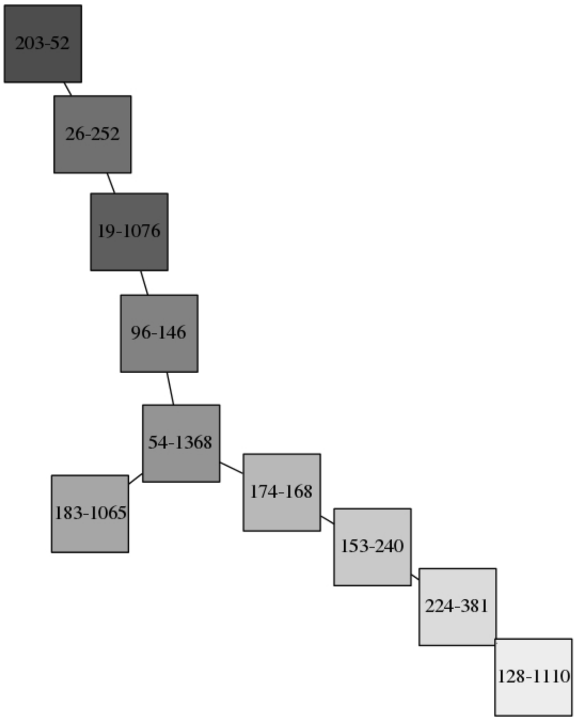


Figure 6.1: Example of a seriation graph, which represents two separate seriation solutions, with one assemblage (54-1368) represented in both subsets. When overlaid, the separate solutions form a graph rather than a single linear ordering. Shading of assemblages reflects temporal information, with light as early and late as dark.

seriation we perform has only a few assemblages and comes out as a linear or near-linear order, seriations will *underdetermine* our hypotheses. Thus, we need to be able to find valid solutions with dozens of assemblages, at least, if not more. It is well known that the basic brute-force seriation problem faces a combinatorial explosion of possibilities, and as described in Chapter 4 breaking the solution into subsets or general graphs only makes the number of possible solutions larger. Lipo and I revisited earlier work by (Kadane, 1971) and added distance minimization as a criterion to our seriation algorithms, allowing much larger sets of assemblages to be ordered than with the frequency criterion, but yielding identical results in nearly all cases. That work is reported in Chapter 5.

These methodological revisions to seriation yield a method whose output—seriation graphs—are comparable to phylogenetic trees in tracing the structure of homology, but doing so with all of the frequency information at our disposal. Seriation graphs thus complement phylogenetic trees at the mesoscopic level of analysis. Each seriation graph documents the "evolutionary chronicle" of how cultural traits were transmitted through a region by teaching and learning of the young, migration, marriage and kinship structure, and trade. The seriation graph is an "observable variable" which we can analyze, cluster, and potentially fit to models which describe the history of cultural transmission at regional scales.

6.3 Representing Hypotheses About Regional Transmission History With Temporal Networks

In order to assess that fit, we need to express our hypotheses about regional history in a quantitative, structured way. We must frame hypotheses about the regional history of cultural transmission which are appropriate given the kind of aggregated, coarse-grained data we are able to possess. For purposes of this study, I focus on one particular kind of empirical case: the archaeological record of nucleated, sedentary communities. In such cases, we tend to have data on artifact class frequencies from "sites" or partitions of sites, sometimes over subsets of the complete interval of occupation, but at the very least, covering the entire occupation duration. At worst, our assemblages thus represent an aggregation over some local population, over some period of time. The frequencies of cultural traits we have to work with are thus coarse-grained at the population level, and time averaged.

This means that we have to model the history of how cultural traits were transmitted at the same level of detail or higher. We can model this in the form of a graph (Diestel, 2010; Harary, 1969). Graph representations have been used for decades in archaeology (e.g., Hunt, 1988; Irwin-Williams, 1977; Mills et al., 2010; Mills, 2017; Terrell, 1977). We can represent the presence of transmission—via any mechanism, whether flows of people by migration or intermarriage, trade, observation and imitation of one's neighbors—simply as a link between two vertices. If we can somehow measure the relative intensity, averaged over time, of the flow or adoption of cultural traits between two communities, we can represent that as a numerical "weight" on the edges of the graph. So far this is standard static graph theory as employed in the growing field of "social network analysis."

Can we use static graphs as models, or do we need a time varying representation? There is good reason to believe that static graphs are not suitable for the type of model proposed here, based on the time scale of the processes we are trying to model and the time scale at which graph structure changes. When the time scale at which the structure of a graph model changes is much slower than the time scale at which a stochastic process on that graph operates ($T_g \ll T_{sp}$), we can consider the graph static and unchanging from the perspective of the process. Conversely, if the connectivity pattern changes much faster than the process we are studying ($T_g \gg T_{sp}$), then we can simply operate with a kind of "average" connectivity pattern (which physicists refer to as an "annealed" graph model). When the structure of a graph and the process we want to study change on similar time scales ($T_g \sim T_{sp}$), then we need to employ a time varying or temporal network to capture the effect of change on the processes we study (e.g., Perra et al., 2012b,a).

In the present case, if we have a single set of artifact class frequencies that are time averaged over the duration of an assemblage, what we can hope to understand is the way that information flowed, by cultural transmission, over the time scales that the durations represent. The graphs we form to describe those flows will thus potentially change at the same rate or slightly slower than our observations. This argues for a time varying or "temporal" graph model, to frame our hypotheses about cultural transmission histories.

Temporal network models add a time dimension to graph models by recognizing that edges may change their weight, or even their presence and absence, and that vertices may go away or arise at various points in time. Temporal networks can be represented in two ways, depending upon the nature of the "contact"' they represent. If an edge represents an instantaneous event (from the perspective of the overall model), we refer to the resulting model as a "contact network". If we model edges as having non-trivial durations, the resulting model is an "interval network." Contact networks are a useful framework for studying individual-level behavior, constructing realistic epidemiological models, or gene expression and protein regulatory networks in cell biology, for example (Holme and Saramäki, 2012). For our purposes in archaeology, interval temporal networks will be most useful, since we need to represent the duration over which each set of class frequency observations spans.

An interval temporal network (ITN) is a graph G with vertex set V, where each vertex v specifies a tuple $(t, \delta t)$, which denotes the time index and duration for which each vertex exists, and an edge set E, where each edge carries a tuple $w, t, \delta t)$, giving the edge weight and the time index and duration over which that edge exists with that weight value. Visualizing this graph is very difficult all at once, since most of the information about the structure of the graph at any point in time is carries in numbers associated with each vertex and edge (Figure 6.2).

Instead, it is convenient both visually and when operating on an ITN computationally to decompose the single model into a sequence of separate graphs. Each graph G_t in the sequence represents one or more change events within the network between times t_i and t_j where i and j represent the union of "change" events in the temporal attributes from the vertex and edge sets. In other words, every addition or loss of an edge or vertex triggers a new G_t in the sequence, and each subgraph in the temporal sequence describes the state of the ITN in terms of vertices and edges over some duration of time where we can observe no change (given the resolution of the data we possess). We must assume that there is an unknown amount of fine-grained change occurring over the interval represented by each subgraph G_t but that variation is not available to us; in this way, the interval temporal

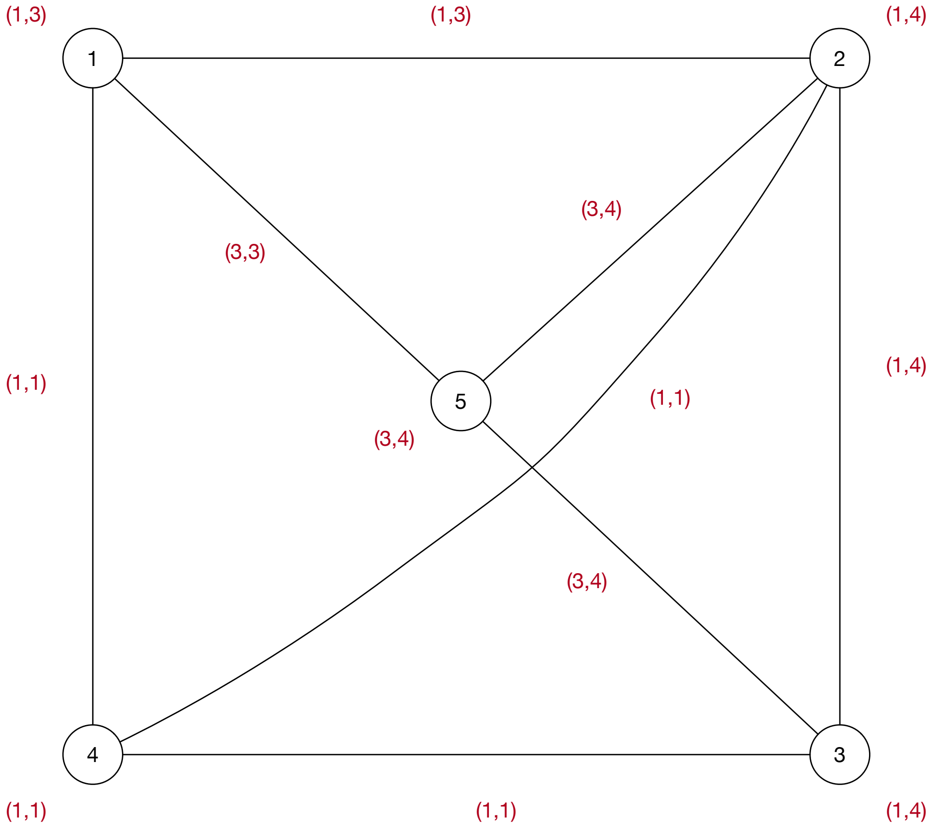


Figure 6.2: Example of an interval temporal network, viewed as a single graph object. In addition to the familiar vertices and edges of a non-temporal graph, each vertex and edge is annotated with the intervals of time over which that object existed. Time in this simple example is arbitrary, and simply represents times at which change events occur.

network representation "naturally" allows us to incorporate the time averaging and coarse-graining of our observed data as part of our observational model. Figure 6.3 displays the same interval temporal network as Figure 6.2, but decomposed into a series of "slices" through time at change points, as vertices go away and a new one arises, with changes in the pattern of connectivity.

Edges in interval temporal networks such as those depicted here may also carry edge weights, which can be used to represent variation in the degree of migration or communication between communities. In simulation, edge weights are transformed into a "migration matrix" or into probabilities of imitating or learning from someone in a different community. Weight variation is the principle mechanism by which a temporal network model would represent "hierarchy" in a regional model, whereby some links and places are more central and important than others. Edge weights are not depicted in Figures 6.2 or 6.3 to make it clear how the temporal aspect of the graph model "works," but they are an integral and important aspect of this modeling framework (for an example of changing edge weights, see the example in Figure 6.4).

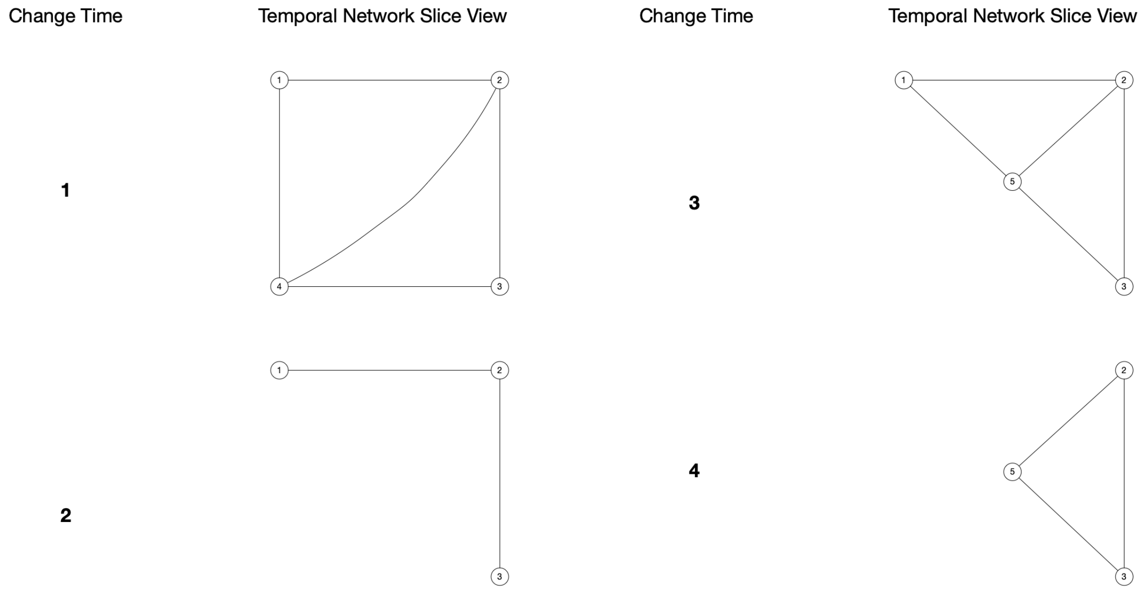


Figure 6.3: Example of an interval temporal network, viewed as a series of "slices" at time indices that represent structural changes. Time in this simple example is arbitrary, and simply represents times at which change events occur.

The interval temporal network is thus a tool for representing networks or interaction patterns not just at a point in time (as in most uses of graph theory for social network analysis) but as those relationships evolve and change. This gives us a way to represent models or hypotheses about how the structure of cultural transmission and regional interaction may have changed, in particular regions or areas.

Such scenarios are necessarily coarse-grained, like our data, and would represent the the flow of cultural traits, whether through the flow of people, material objects, as pure information, by edges in the graph. Vertices represent communities or subpopulations in particular places. The samples of artifacts we obtain archaeologically provide us temporally aggregated samples of cultural variation from those communities. Because that information is time averaged, we represent that duration as the start and end times for the vertex in the ITN.

In our hypotheses about evolutionary history in a region, we can represent different scenarios about how information flows by the way we choose to represent edges in our models. If communities were in contact with most of their neighbors and there is no strong hierarchical pattern, as we might expect in early agricultural communities of the Late Woodland for example, then we could model this with even edge weights. If communities largely interacted with those around themselves, with a smaller number of long-distance links (perhaps for exogamous marriages), we might expect a more lattice-like pattern of "nearest neighbor" edges. The hierarchical structures posited for the "complex" Mississippian communities of eastern North America might imply edge patterns structured such that certain edges had stronger weight than others. These patterns then need to be modeled in the temporal dimension: we can explicitly model the process of small agricultural communities aggregating to form larger and more hierarchical units, and examine the consequences for cultural transmission and the pattern of cultural variation it might leave behind. We can model the process of the development of separate, divergent "traditions" in a region, or the process of several independent "lineages" coalescing to form a larger group.

In the representation developed here, each interval temporal network G as a whole (not its subgraphs) represents a single possible "history" of the presence, absence, and intensity of the sharing of cultural traits between communities over time, as well as the timing and duration of the archaeological assemblages representing those communities. There are many possible histories, and our goal is to find sets of histories which most closely match the empirical data we have concerning cultural trait frequencies in these assemblages. The notion of a "set of histories" is important, because at a detailed level even if we have good knowledge of the vertex set (e.g., all of the Mississippian town sites in the central Mississippi River valley) and their durations of occupation, there may be many possible ITN models which correspond to the similarities and sequences of changes in class frequencies that we observe. Additionally, the vertex set we operate with always represents a sample of the full archaeological record, since our knowledge of that record is incomplete, parts of the record have been fully destroyed by contemporary development, or are otherwise unavailable for study. Thus, we are always examining *equivalence classes* of transmission scenarios. As individual models, the members of an equivalence class are equifinal with each other given the data we possess, and the data we posses underdetermines any distinctions between individual models. The question thus arises: what kind of transmission scenarios can we distinguish using the kind of coarse-grained, time averaged data on artifact class frequencies we typically possess? In this paper I study that issue from a theoretical perspective in order to build a statistical and computational approach to identifying the class of transmission scenarios which fit the data from a case study using ceramic assemblages from the Lower Mississippi River Valley (Section 6.5.2).

6.3.1 Transmission Scenarios Studied

In this study, I employ very coarse-grained transmission scenarios, of the kind just described. The scenarios considered here are simple ones:

- Complete networks: all communities exchange information and individuals with each other;
- Nearest neighbor networks: Interaction is strongly or weakly biased toward nearest neighbors, with small numbers of longer distance links;

- Lineage splitting: a single complete network loses enough links that non-communicating subsets are formed which evolve on their own;

Nearest neighor networks were examined in two variants, to see if the actual spatial arrangement can be distinguished. One model was long and thin ("rectangular nearest neighbor") to mimic communities interacting only with their neighbors up and down a river, for example. A second model was square ("square nearest neighbor") to provide most communities with interacting neighbors in all directions. With an example of a complete network where every community interacted with all others, and a lineage-splitting example where an initial large population splits into two non-interacting lineages, these form the scenarios considered in the present study.

These four scenarios are very basic, and obviously do not encompass the full range of regional histories we might wish to examine for a region as richly complex as the late prehistoric Mississippi River valley. For example, the interaction pattern between mound center and residential sites, and the relative size of mounds underlies a lot of the discussion around social complexity in the Mississippian of the American southeast (Blitz, 2010; Cobb, 2003; Lipo, 2001a,b). An ideal outcome of the line of inquiry pursued in this study would be that large enough sets of assemblages, analyzed into seriation graphs (Lipo et al., 2015), would be comparable to transmission scenarios which represent hierarchical patterns of information flow. This study focuses on the coarse-grained scenarios described here, principally because development of the conceptual and computational methods needed. Future work will need to focus on more precise ways of expressing our archaeological hypotheses in the form of interval temporal networks, and understanding the tradeoffs of different ways of representing something like hierarchy.

For each of the general scenarios described in this section, I constructed interval temporal networks that exhibited that structure. An illustration is shown in Figure 6.4 provides an example of a "decomposed" representation of a "lineage splitting" scenario, where a regional population, with two subpopulations that interact to a limited extent earlier in prehistory, become largely separate in their cultural repertoires over time through loss of direct cultural transmission.

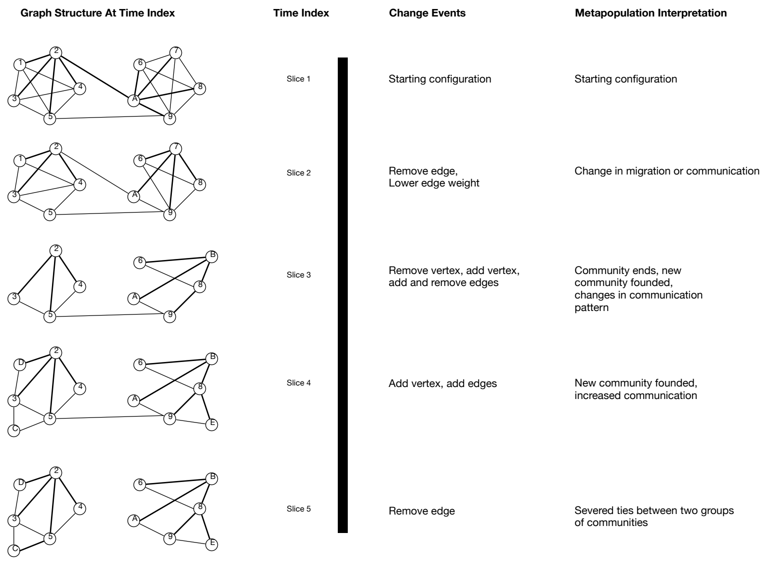


Figure 6.4: Example of an interval temporal network, which can be interpreted as a regional transmission scenario of the lineage-splitting type, with vertices representing communities, edges (with weights indicating relative migration and communication intensity) representing interaction and migration, and changes between subgraphs in the sequence representing changes in patterns of transmission, and the establishment and loss of communities, over time.

In the present study, each of the four scenarios given is represented by a number of randomly generated interval temporal networks, that follow the overall structural pattern of the scenario. Each scenario is constructed as a series of "slices" as depicted in Figure 6.4, with randomly chosen edge "wirings" that follow the needed pattern: complete or nearest neighbor plus infrequent long-distance links. In the case of lineage coalescent or splitting scenarios, the vertices that belong to the non-communicating components are randomly assigned and then links broken either starting with or ending at a designated time step. This process of building replicas of interval temporal networks allows us to randomize over the details of network structure, to ensure that results are not idiosyncratic.[3]

6.4 Methods

6.4.1 Study Design

The goal in this study is to determine whether it is possible, given only the seriation graphs derived from simulations of cultural transmission in the four regional transmission histories described in the previous section, to quantitatively identify the proper data generating model. Specifially, our goal is to determine whether the structure and topology of the resulting seriation graphs is diagnostic of the transmission scenario used to generate the data. This test is thus one that seeks to determine *equifinality between transmission scenarios*, and whether our observable tools—seriation graphs—have discriminatory power. The data generating half of the study proceeds as shown in Figure 6.5.

Given a large sample of simulated seriations, equifinality among theoretical transmission scenarios is evaluated using a machine learning classifer (the same approach taken in Chapter 3). In order to train a classifier model, we first transform the seriation graphs into a numerical representation using their Laplacian spectrum (see Section 6.4.2. The data are randomly split into a training set, and a hold-out test set to evaluate equifinality in terms of classifier accuracy. Accuracy will be evaluated using the confusion matrix of correct and erroneous predictions, to determine which pairs of transmission histories are distinguishable, and which if any, display equifinality at the theoretical level. Figure 6.6 depicts this workflow.

---

[3] All of the experiments described in this study were created with the software located at `https://github.com/mmadsen/seriationct`. The subdirectory **graphs** contains a number of standalone progams written in Python 2.7. Each is a "generator" for transmission scenarios as interval temporal networks, represented as a number of slices stored in GML (Graph Modeling Language) format.

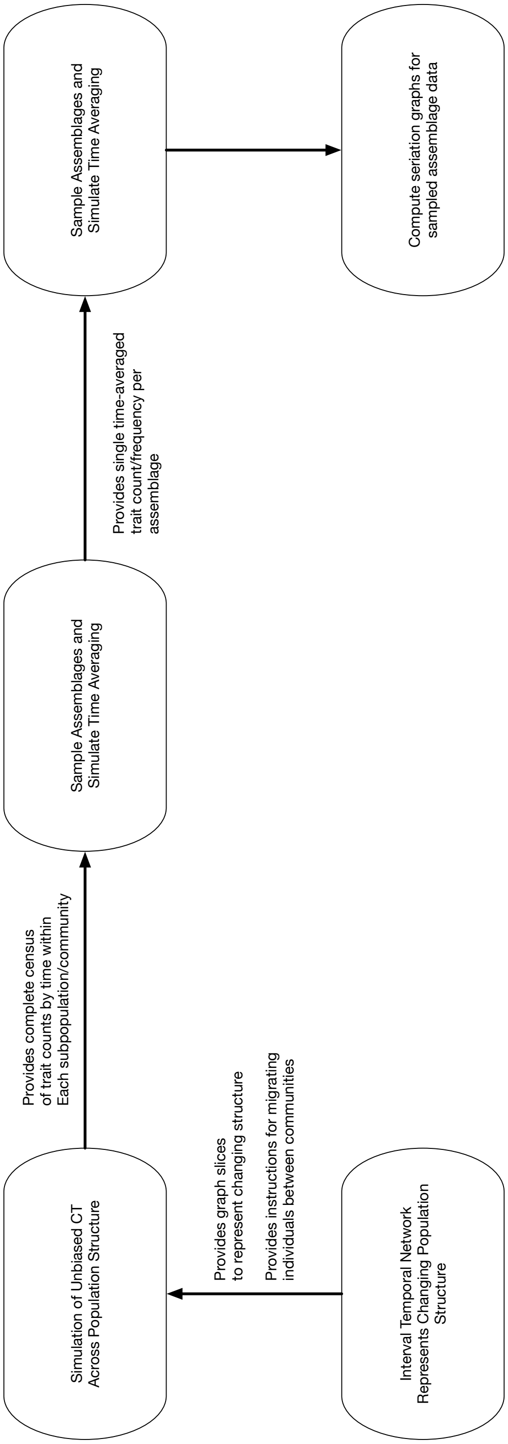


Figure 6.5: Flowchart for simulated data generation from each transmission scenario. Samples of transmitted traits from each community in the transmission scenario are then aggregated in realistic ways and then the resulting trait frequencies are seriated using the IDSS seriation algorithm (Lipo et al., 2015), to produce seriation graphs as output.

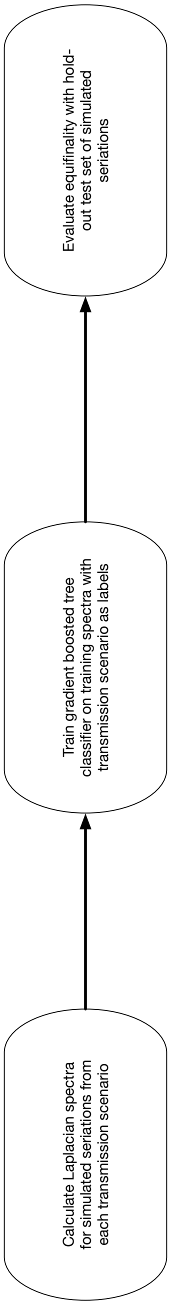
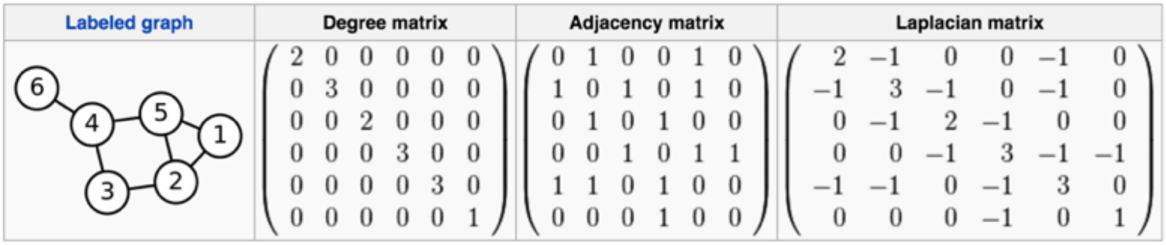


Figure 6.6: Flowchart for evaluating equifinality among the four candidate transmission scenarios. Simulated seriation graphs from each model are rendered numerically using their Laplacian eigenvalue spectrum, and these eigenvalues are employed as predictor variables to train a gradient boosted tree classifier, which is then tested for predictive accuracy (and the lack of equifinality) on a hold-out set of seriation graphs.

6.4.2 Quantifying The Structure of Seriation Solution Graphs

In order to make comparisons between seriation graphs, and determine our ability to predict which transmission scenario generated a particular graph, we need to quantify something about the structure and topology of the graphs. There are many types of graph metrics, but many are applicable to general graphs which contain circuits and loops (Chebotarev, 2013; Diestel, 2010). The seriation graphs being constructed here are trees and have a single connected component by construction, rendering many classical graph metrics useless. Instead, we turn to algebraic and spectral graph theory, which characterize the properties of trees and general graphs using the numerical properties or "spectra" of the various matrices associated with a graph (Banerjee and Jost, 2008; Beineke et al., 2004; Chung and Graham, 1997; Godsil and Royle, 2001).



Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2&0&0&0&0&0\\ 0&3&0&0&0&0\\ 0&0&2&0&0&0\\ 0&0&0&3&0&0\\ 0&0&0&0&3&0\\ 0&0&0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&0&1&0\\ 1&0&1&0&1&0\\ 0&1&0&1&0&0\\ 0&0&1&0&1&1\\ 1&1&0&1&0&0\\ 0&0&0&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 2&-1&0&0&-1&0\\ -1&3&-1&0&-1&0\\ 0&-1&2&-1&0&0\\ 0&0&-1&3&-1&-1\\ -1&-1&0&-1&3&0\\ 0&0&0&-1&0&1 \end{pmatrix}$

Figure 6.7: Simple example of calculating the Laplacian matrix for a small general graph example. The basic Laplacian matrix is simply the element-wise difference between the degree and adjacency matrices.

The "structure" of seriation graphs is captured in two general characteristics: the number of branches and neighbors that assemblages possess, and the distances between assemblages. The more a seriation is long and linear, the longer the average "distance" along the tree two assemblages chosen at random will be, and the more "branched" and reticulate the seriation is, the more average distances will decline, even as average vertex degree increases. These intuitions are captured neatly in the *Laplacian* matrix of a graph, which is defined as the difference between the "degree matrix" and the "adjacency matrix", as demonstrated for a simple example in Figure 6.7. The degree matrix is a diagonal matrix where the entries along the diagonal record the number of edges attached to each vertex (i.e., vertex degree). The adjacency matrix (for an unweighted, undirected graph, with no self-loops) has zeroes on the diagonal, while the off-diagonal elements contain 1 if an edge is present between two vertices, and zero otherwise. The Laplacian matrix thus encodes information about a graph's connectivity. The "spectrum" is then simply the eigenvalues (and their multiplicities) of the Laplacian matrix.

Certain classes of graphs have spectra that contains every bit of information possible about the graph, and are thus "determined" by their spectra (e.g., finite star-like trees and the complete graphs). In most other cases, spectra define equivalence classes of graphs with very similar structure. Two graphs with the same spectrum of eigenvalues (and multiplicities) are *cospectral* and share the same connectivity pattern even if they are not identical or isomorphic. Further, even if two graphs are not isomorphic or cospectral, their eigenvalue spectra will be more similar to the degree they share common connectivity structures. These properties make the Laplacian spectrum a compact numerical way to capture the structure of a graph and compare it to many other graphs. Thus, in examining the degree to which we can predict the transmission scenario which generated a given seriation graph, what we are really asking is whether we can construct a clean partition or clustering of graph spectra arising from the four transmission scenarios given in Section 6.3.1. In this study, I employed the Python *NetworkX* library to calculate Laplacian matrices and spectra for seriation graphs.

6.4.3 Simulation of Cultural Transmission on Interval Temporal Networks

Simulation of cultural transmission within the interval temporal networks which represent the transmission scenarios listed in Section 6.3.1 employ the standard unbiased copying model, since previous work demonstrates that it is difficult or impossible to detect more detailed transmission biases from the kind of coarse-grained data this study targets (see Chapter 2 for more details on the standard Wright-Fisher representation of unbiased cultural transmission). The simulations were performed with a custom Python library, *SeriationCT* written by the author, which adapts the *SimuPOP* forward-time pouplation genetics simulation package to employ interval temporal networks as population structures, and performs the necessary sampling and temporal aggregation to simulate time averaged archaeological assemblages.[4]

Parameter	Value or Interval
Innovation rate (in θ scaled units)	$[0.00005, 0.0001]$
Simulation length	10000 steps
Sample fraction	0.5
Migration fraction (between communities)	[0.05, 0.1]
Individual population size	250
Number of trait dimensions (loci)	3
Initial traits per dimension	5

Table 6.1: Parameters for simulation runs across the four models studied. Intervals are treated as prior distributions, and each simulation run is assigned values derived from a uniform random sample on the interval indicated. Single values are applied to every simulation run, and represent a point prior.)

Specifically, simulations operate in the following manner:

- Each individual is characterized by 3 dimensions of variation, which start out having 5 initial traits in the population;
- Innovation follows the standard "infinite alleles" model of mutation/innovation;
- At each time step, individuals have a probability of copying a trait from a randomly chosen individual or keeping their existing trait set. Copying occurs within the local population only;

---

[4] SeriationCT is open source software, available on Github at `https://github.com/mmadsen/seriationct`. SimuPOP was written by Bo Peng, and is available on Github at `https://github.com/BoPeng/simuPOP`, and documented in his book (Peng et al., 2012).

- At each time step, there is a probability of individuals moving between communities. Movement is governed by a migration rate probability and strictly follows the edge pattern of the interval temporal network model in force for that simulation run, at that time step;
- At specific time steps, the interval temporal network defines changes to the connections between communities, and that some communities may go away, or new communities coming to be;
- New communities, should they arise, are seeded with individuals from the neighboring communities that possess edges in the network to the new community, to create a consistent sample of cultural continuity given the transmission scenario being simulated;
- At each time step, the frequencies of traits at each locus or dimension are tabulated, as are the cross-tabulation of "classes" formed by intersecting the loci;
- Class frequencies are aggregated over the entire duration that a community (vertex) exists in the interval temporal network model, to form time averaged "assemblages" for seriation analysis;
- Before performing seriations, assemblages were sampled from all of those available in a given simulation run, over the entire time course. This simulates our partial view of any given regional archaeological record given what we have collected or excavated compared to the totality of the record.

Simulations were run with the parameters given in Table 6.1, with parameter ranges treated as "prior distributions" and sampled uniformly. This allows the simulation results to be treated as a proper Monte Carlo sampling of the space formed by the priors, for approximate Bayesian computation and other analytic methods. 500 simulations were performed for each scenario.

Seriations were performed using the Python *IDSS* seriation package, written by Carl Lipo and the author for our 2015 paper on iterative deterministic seriation solutions, which introduced seriation graphs (Lipo et al., 2015).[5] Raw simulation results, all parameter choices and shell scripts to run each stage of the ITN creation, simulation, resampling, and seriation pipeline, are available in a Github repository (`https://github.com/mmadsen/experiment-seriation-classification`).[6] Not every random sampling of assemblages from a simulation run yielded data which could be seriated due to the vagaries of sampling. The final number of valid solutions across the four scenarios was 1946.

6.4.4 Classifier Training and Accuracy Evaluation

This study employed the same gradient boosted tree classifier model employed in my previous work on equfinality and cultural transmission (see Chapter 3). The simulated seriation graphs were randomly split into a 90% training set and a 10% hold-out test set, within each transmission scenario. Each seriation graph was transformed into its Laplacian eigenvalue spectrum using a utility library of *scikit-learn* compatible statistical functions maintained on Github at `https://github.com/mmadsen/sklearn-mmadsen`. The resulting eigenvalue spectra became the input predictor variables for the gradient boosted classifier with numerical labels denoting each of the four transmission scenarios. For this study I employed the gradient boosting implementation in the Python *scikit-learn* package. Optimal hyperparameters were found by optimization grid search with 3-fold cross validation, with learning rates ranging from 5.0 down to 0.01, and the number of trees from 10 to 500. Optimal performance on the training set was achieved with learning rate 0.05 and 500 trees. Final results were obtained by predicting the transmission scenario for each seriation graph in the hold-out test set in the same manner, and calculating the overall accuracy, F1 score, precision, and recall for the

---

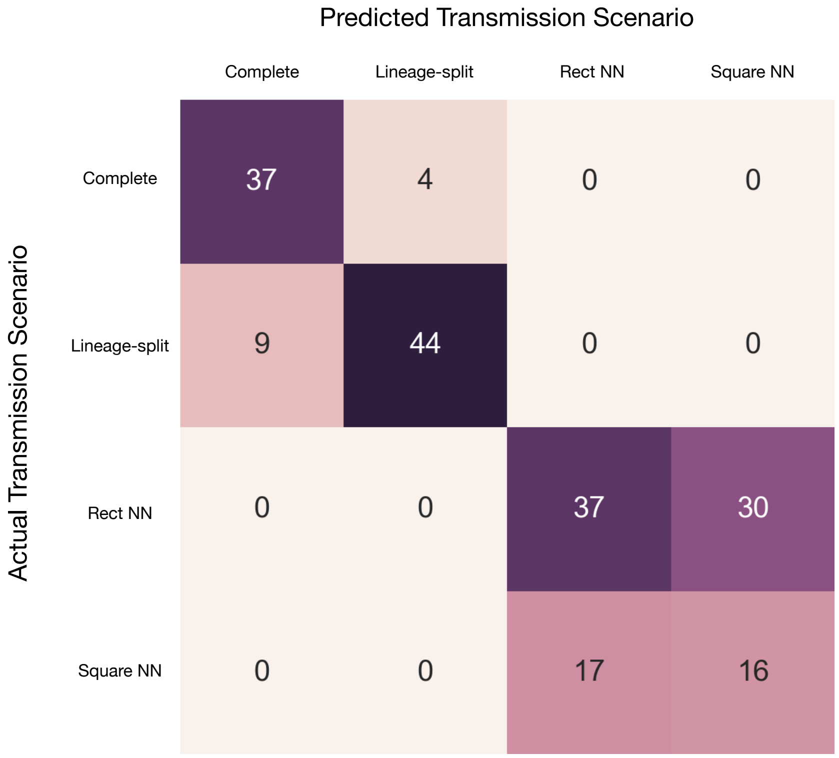
[5] The IDSS package is open-source software, freely available at `https://github.com/clipo/idss-seriation`.

[6] Additional experiments and prototypes for performing this kind of classification of transmission scenarios are located in the repository `https://github.com/mmadsen/experiment-networkmkodel-identification-diss`, and discussed online at `http://notebook.madsenlab.org`.

resulting confusion matrix. The test set comprised 194 seriation graphs spread across the four transmission scenarios.

6.5 Results

6.5.1 Equifinality Analysis of Transmission Scenarios with Simulated Data



	Complete	Lineage-split	Rect NN	Square NN
Complete	37	4	0	0
Lineage-split	9	44	0	0
Rect NN	0	0	37	30
Square NN	0	0	17	16

Figure 6.8: Classifier results for four regional cultural transmission history scenarios, using Laplacian spectra to determine whether seriation graphs from simulations of cultural transmission under each scenario are cleanly separable. Results refer to evaluation of the 10% hold-out test set of 194 seriation graphs.

The results of equifinality analysis between the four transmission scenarios are given in Figure 6.8. The confusion matrix has correct assignments on the diagonal: the predicted transmission scenario matches the actual transmission scenario under which a particular simulation result was generated. In general, accuracy is only 69.1%, but this is driven down mainly by the inability to distinguish between our two variants of nearest-neighbor transmission scenarios. The F1 scores for cases actually coming from complete networks was 0.85 and 0.87 for lineage-split scenarios. It is also quite clear that one can distinguish between complete networks, lineage-splits, and nearest-neighbor scenarios *in general*, given no errors.

In these results, it seems less clear that the spatial "shape" of the contact network had a distinguishable effect on the Laplacian spectra of the seriation graphs. This could be due to the interaction between quality of spatial sampling and overall sample size in seriations. Although the assemblages selected to seriate (from the overall set present in simulations) were stratified by space to ensure that the entire space was sampled, with 20 assemblages also spread out in time, there may not have been sufficient numbers of data points to provide detail on the different spatial configurations. We should expect, in fact, that discrimination between fairly "close" transmission scenarios will require larger sample sizes than discriminating between scenarios which are structurally quite different.

In general, these results are encouraging for the approach studied here. Seriation graphs may, with appropriate choice of scenarios to contrast, be capable of being statistically identified as to the equivalence class of regional cultural transmission scenarios that drove the empirical cases we see in prehistory.

6.5.2 Analysis of Lower Mississippi River Valley Ceramic Data

Given the trained classifier model, we can analyze empirical examples and determine the degree to which the model presents a confident prediction for the transmission scenario that might correspond to our data from prehistory. I calculated the Laplacian spectrum for a frequency seriation result from 20 ceramic assemblages from Phillips, Ford, and Griffin's (1951) study of the Lower Mississsippi

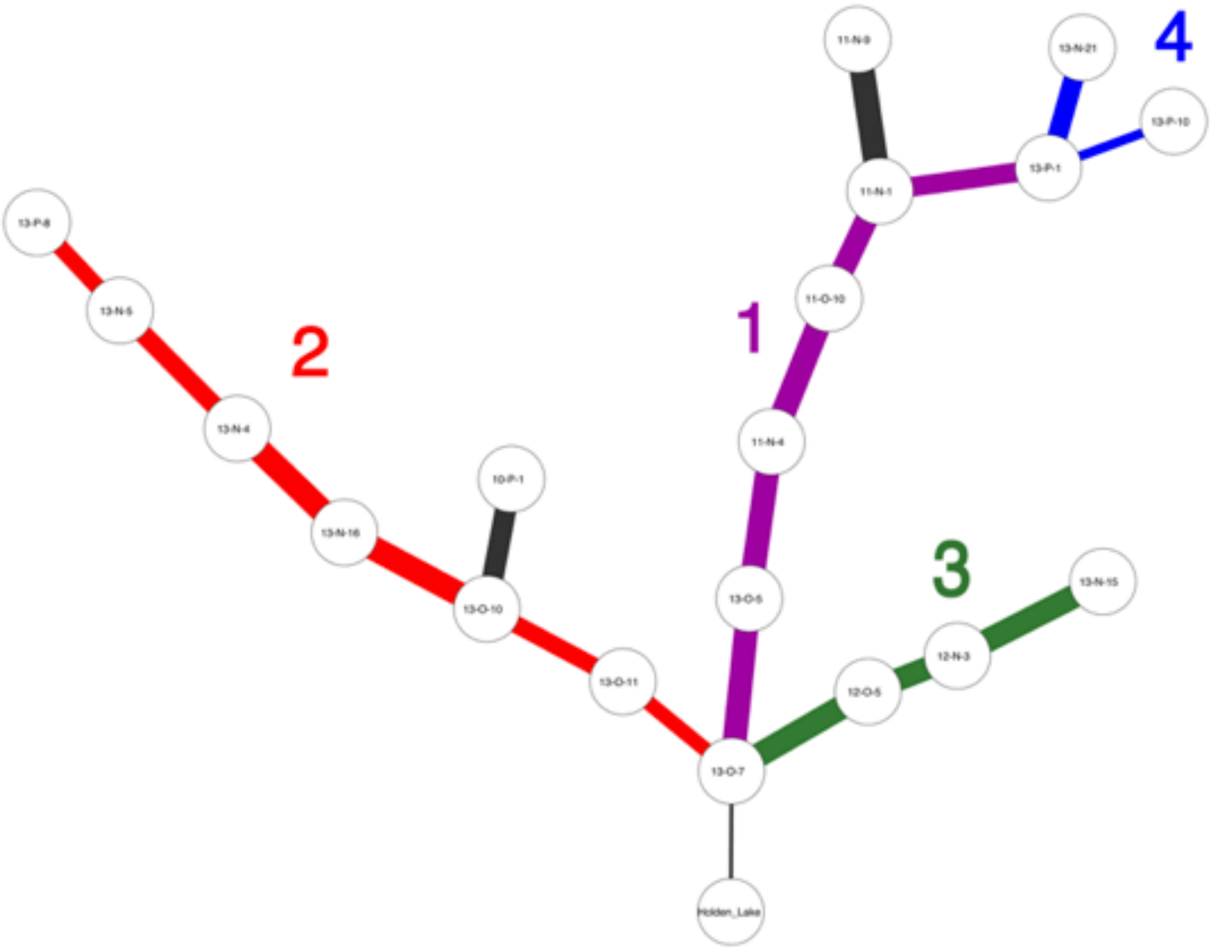


Figure 6.9: Seriation graph for PFG (1951) ceramic assemblages in the Lower Mississippi River Valley, as analyzed by Lipo (2001a) and re-analyzed by Lipo et al. (2015). Colors and numbers correspond to spatial clusters as mapped in Figure 6.10.

River Valley, as augmented and reanalyzed by Carl Lipo (Lipo, 2001b). The resulting seriation graph is shown in Figure 6.9. There is a clear branching structure, which Lipo interpreted in his dissertation work as reflecting an early assemblage—Holden Lake—being ancestral to other early Mississippian sites in the region, but giving rise to several somewhat independent regional ceramic traditions. There is also clear spatial structure to this results, as shown in Figure 6.10.

When we use the trained classifier with the Laplacian spectrum of this seriation graph, the predicted transmission scenario is unsurprisingly the lineage split model, with probability 0.93. The complete graph scored 0.06 probability, with the remaining 0.01 split between both nearest neighbor models. This example is suggestive, even though the number of possible scenarios is small. But it demonstrates the complete method by which empirical data would be processed to identify matches with possible transmission scenarios.

6.6 Discussion

O'Hara (1988) usefully distinguished between the evolutionary *chronicle*, which is simply the facts what happened when, from evolutionary *history*, which is a narrative of how and why things happened the way they did. At the macroevolutionary scale, phylogenetic methods produce tree structure which abstractly depict the chronicle of homologous relationships between taxa. It is clear that the trees are data, not evolutionary history itself: we still need to posit hypotheses about the processes and events which resulted in the observed trees. This applies whether we are talking about genetic evolution with species as taxa, or cultural evolution with samples of cultural variation or individual artifacts as the taxa.

As we develop methods at the mesoscopic level, making use of class frequency data to attempt to study evolutionary pheneomena within regions over time scales of decades or a century, we should keep the distinction between chronicle and history firmly in mind. In this study, I have attempted to develop a computational method for determining the statistical fit between seriation graphs (the *chronicle*) and interval temporal network models which formalize the regional history of cultural transmission (the *history*). That method employs a modified seriation method which produces graphs rather than linear orderings if the data require it; the branchings in seriation graphs indicate relationships where the history of cultural traits varies through space as well as time. I also introduced the "interval temporal network" as the natural way to formalize our hypotheses about the mesoscopic history of cultural transmission within and among sedentary, nucleated populations, as those populations arise, grow, and eventually go away for various reasons. Machine learning classifiers seem able to resolve these hypotheses (in some cases) without equifinality, although there is much to be investigated within this area of research.

Given the coarseness of the transmission scenarios studied here, it is not clear yet if we can determine from seriation graphs whether "hierarchy" is identifiable, although this is an important next step for the study of Late Prehistoric and Late Holocene populations in many regions. There is some evidence that spatial information may only be weakly encoded in seriation graphs, since we could not distinguish nearest neighbor hypotheses even in simulated data with different spatial configurations. But more controlled study about the degree to which seriations encode space and with what resolution is clearly necessary; all of our knowledge of this is anecdotal given the "same local area" criterion long used for seriation analysis.

Despite the potential limitations of our ability to resolve differences among theoretical models, evolutionary archaeologists will benefit from further work along these lines to develop a rigorous understanding of our ability to fit theoretical models to data at a variety of spatiotemporal scales. Phylogenetic methods have proven their worth but answer questions at particular scales and levels of detail; methods for answering more detailed questions (while still operating with coarse-grained, time averaged data) have been lacking. Seriation, properly construed as a method for mapping homology using all of the frequency data at our disposal, provides one such tool.

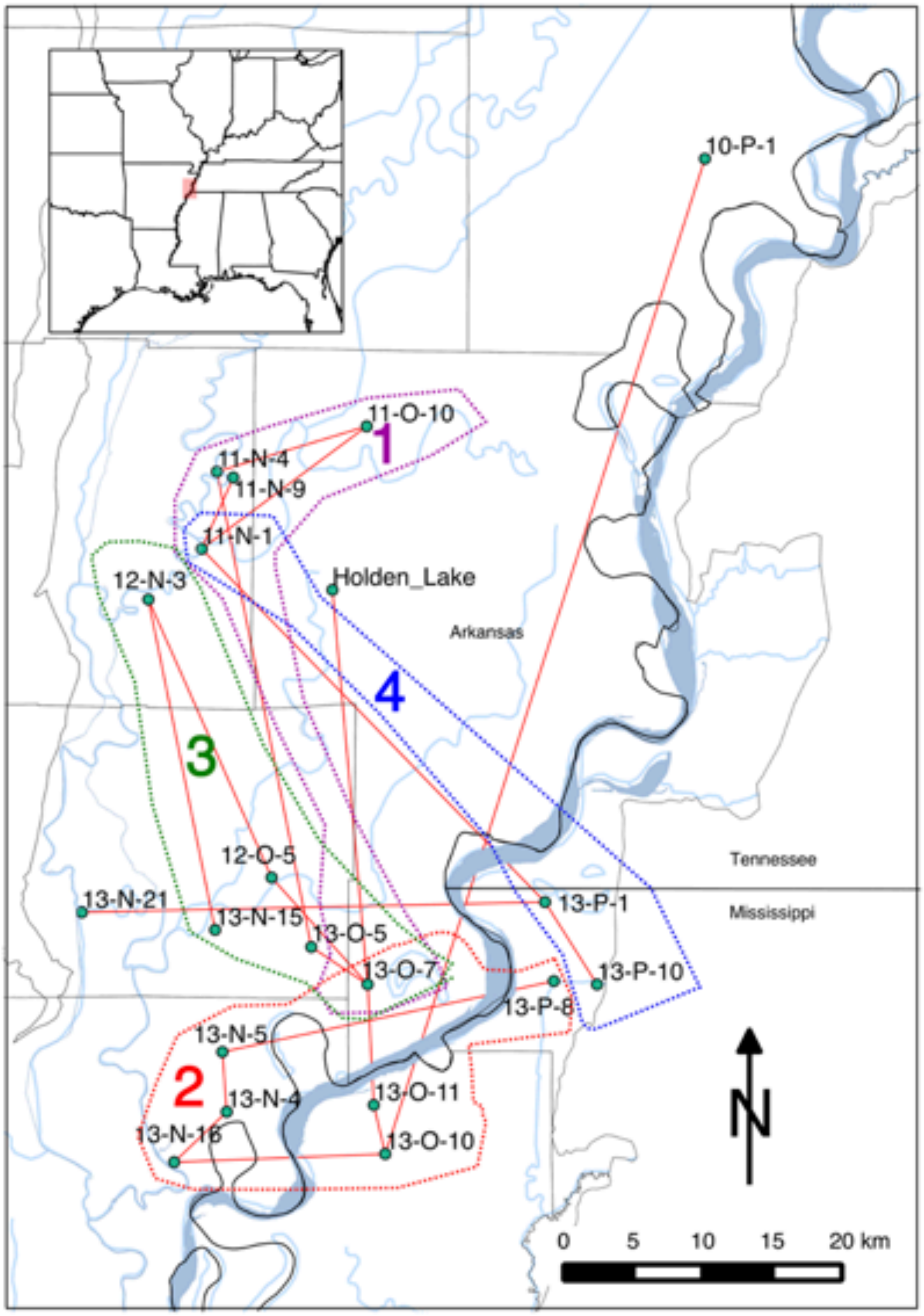


Figure 6.10: Spatial distribution of ceramic assemblages from the Lower Mississippi River Valley, as analyzed by Lipo (2001a), corresponding to the seriation graph in Figure 6.9.

CHAPTER 7

Behavioral Modernity and the Cultural Transmission of Structured Information: The Semantic Axelrod Model

Abstract Cultural transmission models are coming to the fore in explaining increases in the Paleolithic toolkit richness and diversity. During the later Paleolithic, technologies increase not only in terms of diversity but also in their complexity and interdependence. As Mesoudi and O'Brien (2008b) have shown, selection broadly favors social learning of information that is hierarchical and structured. We believe that teaching provides the necessary scaffolding for transmission of more complex cultural traits. Here, we introduce an extension of the Axelrod (1997) model of cultural differentiation in which traits have prerequisite relationships, and where social learning is dependent upon the ordering of those prerequisites. We examine the resulting structure of cultural repertoires as learning environments range from largely unstructured imitation, to structured teaching of necessary prerequisites, and we find that in combination with individual learning and innovation,

high probabilities of teaching prerequisites leads to richer cultural repertoires. Our results point to ways in which we can build more comprehensive explanations of the archaeological record of the Paleolithic as well as other cases of technological change..

Source Reprinted by permission from Springer from: *Learning Strategies and Cultural Evolution During The Paleolithic*, edited by Alex Mesoudi and Kenichi Aoki, 2015, in the series *Replacement of Neanderthals By Modern Humans*, Springer. Co-authored with Carl P. Lipo.

7.1 Introduction

Although humans and our hominid ancestors have been cultural animals throughout our evolutionary history, an important change occurred in our lineage during the Middle and Upper Paleolithic. For millennia our ancestors manufactured relatively small toolkits and their material culture was remarkably similar across continental distances and over many generations. Beginning in the Middle Paleolithic and continuing through the Upper Paleolithic, the archaeological record reflects an explosion in our cultural repertoire. Over tens of thousands of years, artifactual toolkits shift from sets of relatively few objects with multiple uses to large collections of functionally-specialized tools that, employed increasingly complex technologies and that were manufactured from an enriched range of materials. The changes in artifacts suggest that human solutions to the problems of everyday life became regionalized and differentiated. Further, the economic basis of our lives began to broaden and also, in many areas, to become specialized (Bar-Yosef, 2002; d'Errico and Stringer, 2011; Straus, 2005).

While early researchers believed that the Upper Paleolithic resulted from a singular "revolution" in human evolution leading to behaviorally modern homo sapiens, this view is held by a minority of paleoanthropologists and archaeologists today (e.g., Klein, 2009). Careful examination of the

Middle Paleolithic archaeological record especially in Africa and the Near East suggests that this change in behavior did not occur as a single distinct event, instead occurring over a long period of time since much of the enriched material culture we later characterize as the "Upper Paleolithic" had precursors. In addition, this change now appears to be patchy and fitful, with modern features appearing and frequently being lost again (Bouzouggar et al., 2007; d'Errico and Henshilwood, 2007; d'Errico and Stringer, 2011; Straus, 2005; McBrearty and Brooks, 2000; McBrearty, 2007). Nor does behavioral modernity map neatly to biological taxa and their movements, given that evidence for the precursors of fully modern behavior is abundant in deposits associated with Neaderthals in addition to modern *Homo sapiens* (Villa and Roebroeks, 2014).

The "learning hypothesis" studied in this series of volumes makes the plausible claim that behavioral modernity is the product of cumulative changes in the way cultural information was acquired and retained across generations (Nishiaki et al., 2013), thus providing a potential explanation for the slow evolution of "modern" features, its patchiness in space and time, and the lack of a neat mapping between hominin taxonomy and material culture. In short, according to the learning hypothesis, behavioral modernity arose through a change or changes in the way social learning operated within hominin groups, with those groups adopting richer modes of cultural learning surviving and spreading compared to those who retained simpler forms of social learning.

Within the umbrella of the learning hypothesis, there are many ways in which social learning and thus intergenerational cultural transmission could have changed, and an increasing amount of research is focused upon formulating and testing different models. One class of studies is focused upon factors exogeneous to the learning or imitation process itself. Shennan (2000; 2001) proposed that population size has a powerful effect on diversity within cultural transmission processes, which Henrich showed in the case of toolkit element loss during a Tasmanian population bottleneck (Henrich, 2004). In a similar line of reasoning, Kuhn (2013) argues that low population size and density put Neanderthals in a situation where innovations spread slowly and ultimately led to their demise relative to modern humans. Furthermore, a growing set of experimental studies clearly show a relationship between accumulation of complex cultural traits and the number of cultural "models" from whom individuals can learn (Muthukrishna et al., 2014; Derex et al., 2013; Kempe and Mesoudi, 2014). Not all studies have shown a strong association between population size and cultural diversity, however. Collard and colleagues, find little association in a linked series of comparative studies (Collard et al., 2011, 2013a,b,c). Finally, in his analysis of the overall evolutionary rate, Aoki (2013) found that innovation rates were more important than population size to determining the rate of evolution in a population.

To us, this body of work indicates that while population size is an important parameter in mathematical models, it may be better understood as a second-order effect in the real world, interacting with a myriad of other factors and thus often dominated by those factors. Another important factor is the structure of bands or demes into larger regional metapopulations. Network topology, for example, is known to have a substantial effect upon contagion or diffusion processes (e.g., Castellano et al., 2009; Smilkov and Kocarev, 2012). Thus, it is likely that regional structure has critical effects on the outcomes we can expect from a single social "learning rule." Along these lines, Premo (2012) has examined whether metapopulation dynamics that include local extinction and recolonization might provide an improved account for the retention and expansion of diversity.

A second group of studies has focused upon endogeneous changes to social learning processes. Many authors in this volume series, for example, have looked at aspects of the way individuals learn skills and acquire information (Aoki, 2013; Nishiaki et al., 2013). We know that learning and teaching styles vary across human groups, and formal modeling efforts are beginning to make clear that such variation has evolutionary consequences that might lead to a rapid expansion of the human cultural repertoire (Nakahashi, 2013). Those populations which increased the amount or effectiveness of teaching would have a fitness advantage over those who relied upon imitation and "natural pedagogy" in passing along technological and foraging knowledge (Csibra and Gergely, 2011; Fogarty et al., 2011; Terashima, 2013). Demography and population structure would then play an important role in reinforcing the fitness differences which different learning strategies would create, as pointed out by Kuhn (2013).

Ultimately, a full "learning explanation" for behavioral modernity will be multifacted, including demographic and spatial changes as well as changes to the mechanisms of social learning and technological innovation themselves. Sterelny (2012, p.61) sums up this kind of multifactorial approach to behavioral modernity well:

> …the cultural learning characteristic of the Upper Paleolithic transition and later periods of human culture—social transmission with both a large bandwidth and sufficient accuracy for incremental improvement—requires individual cognitive adaptations for cultural learning, highly structured learning environments, and population structures that both buffer existing resources effectively and support enough specialization to generate a supply of innovation.

In research designed to explore how the structure of a learning environment affects the results of social learning, Creanza and colleagues (2013), Aoki (2013), Nakahashi (2013), and Castro and colleagues (2014) developed models that examine how explicit teaching (as opposed to simple imitation) affects the overall evolutionary rate or cultural diversity in a population. Castro et al., for example, find that cumulative cultural transmission requires active teaching in order to achieve fidelity across generations. Our work in this chapter follows these authors, focusing on the nature of transmitted information itself and the effects of teaching upon the richness of structured technological knowledge.

In particular, we suggest that when knowledge is structured such that skills and information must be learned in sequences, high fidelity learning environments are critical to evolving ever-richer cultural repertoires, of the type seen in behaviorally modern assemblages. To formalize this idea, we construct a model which:

- Represents cultural traits as hierarchically structured, in order to study increases in complexity,
- Has a learning rule sensitive to the order in which cultural traits are acquired, with multiple levels of fidelity, and
- Has a mechanism (such as homophily) that allows cultural differentiation endogeneous to the model.

As we alter the "learning environment" in our models from less to more frequent teaching of traits and their prerequisites, we expect to see greater diversity, larger structured sets of traits persisting in the population, and greater differentiation of the population into "different" cultural configurations. We also expect that individual innovation, independent of the social learning context, will play a role in the accumulation of cultural complexity by allowing a population to explore increasingly large spaces of technological design possibilities; this expectation is concordant with Aoki's (2013) result in Volume I of this series.

In this chapter, we introduce a simulation model which combines a hierarchical trait space capable of expressing dependencies or semantic relationships between skills and information (Mesoudi and O'Brien, 2008b), and a modified version of Robert Axelrod's (1997) homophilic social learning model which allows us to examine the conditions under which evolution in a hierarchical design space leads to cultural differentiation. After describing the model, we study its dynamics and provide an initial assessment of its suitability for studying the onset of behavioral modernity in the later

Paleolithic. Models like this begin to move beyond diffusion dynamics, bringing the actual meaning and relations of traits into the modeling process. Hence, we call these "semantic Axelrod" models, and believe that such models form a platform for formalizing the type of multi-factor hypotheses necessary to examine major transitions in human evolution, such as "behavioral modernity."

7.2 The Semantic Axelrod Model for Trait Prerequisites

Much of our technical knowledge, whether of stone tool manufacture, throwing clay pots, or computer repair, is built from simple tasks, bits of background knowledge, and step-by-step procedures (Neff, 1992; Schiffer, 1987). These pieces of cultural information are not simply a set of alternative options, which can be mixed and matched in any combination. Instead, there are dependencies and relationships between items which affect how skills and information are learned and passed on between individuals. Some items will be related in time, as steps in a process. Others will be related by subsumption: arrowheads are a subclass of bifacial stone tools, and require many of the same production techniques as bifaces used in other projectiles. Still others will be related as sets of alternatives: choices of surface treatment for a given ceramic paste, given the firing regime selected, for example. To date, most archaeological models of tool production have focused upon temporal relations in the construction of an artifact, as in "sequence models" or "chaîne opératoire," but it is important to remember that other representations are possible, including trees and more general graphs to capture relations of use, reworking, or discard (Bamforth and Finlay, 2008; Bleed, 2008; Ferguson, 2008; Högberg, 2008; Bleed, 2001, 2002; Schiffer, 1987; Stout, 2002).

Given conscious reflection, we describe and organize our knowledge and skills in many ways, but it is common (especially while learning a new skill) to think of a complex process as a "script" or "recipe" (Schank and Abelson, 1977). Experts in a task or field may not represent their knowledge this way, having internalized such structures below the conscious level. Experts will often know more than one way to accomplish any given goal, and be able to repurpose and recombine methods and tools, as opposed to the simpler, more linear or tree-based recipes of the novice or student (e.g., Bleed, 2008, 2002; Stout, 2002). Nevertheless, it is common to teach or learn new information and skills in a stepwise manner.

In this chapter, we focus not on the execution steps of a recipe (and thus not on sequence models), but the relations between skills and information *during the learning process*. In specific, we focus upon the *prerequisite* relationships that exist between cultural traits, since the ordered dependencies between skills and information form one of the structures within social learning occurs during development (and into adulthood). Some pieces of information or skills must be in place before a person can effectively learn or practice others. Examples from our own childhoods abound: one needed to understand addition and subtraction and multiplication before learning long division; in order to make soup, we need to understand how to simmer rather than boil, how to chop and slice, what ingredients might be combined, and so on. The fact that knowledge and skills build upon one another make prerequisite relations between cultural traits ubiquitous. In this chapter, we represent prerequisite relations as trees in the graph-theoretic sense (Diestel, 2010), replacing the "nominal scale" structure of "locus/allele" models or paradigmatic classifications and some typologies (Dunnell, 1971), but we emphasize that the tree models we discuss here are still classifications and thus analytic tools, designed to allow us to measure variation in the archaeological record, not reconstruct emic models of Paleolithic technologies.

Our model also requires a way of representing a changing learning environment, in ways that create higher fidelity and greater possibility for building cumulative knowledge. In real learning environments, there are many possibilities, but deliberate teaching and apprentice learning are repeatedly seen across human groups as ways that naive individuals can reliably learn the complex skills and information needed for foraging, artifact production and maintenance, and navigating an increasingly rich social world. The point of structuring the learning environment with teaching and/or apprenticeship is to give the learner skilled models to imitate, shortcut trial and error when acquiring a skill, provide a reference for needed information, and to guide individuals to put their information and skills together into appropriate sequences to accomplish an overall goal. Apprenticeship and formalized teaching provide a social learning "scaffold," helping to lower the amount of individual trial and error learning needed to master a body of material (Wimsatt and Griesemer, 2007; Wimsatt, 2007).

Within a standard discrete-time simulation model of a social learning process, we can model this type of learning environment with the following modifications:

1. Represent the order in which skills and information need to be acquired as a series of trees, with vertices representing traits (either a skill or piece of information), and edges the prerequisite relations between them.
2. Disallowing individuals the ability to copy traits from a cultural model for which they do not have necessary background or prerequisites, given the relations in the applicable tree model.
3. Creating a probability that individuals, if disallowed a trait, can be taught one of the needed prerequisites instead by that cultural model, leading to the potential accumulation of fuller knowledge and skills over time.

By changing the probability that individuals learn a missing prerequisite trait, we can "tune" the learning environment. Low probabilities might correspond, for example, to a learning environment where individuals can observe others executing a production step, but are given little or no instruction or guidance on what they need to know in order to successfully master it. High probabilities of learning prerequisites would correspond, on the other hand, to environments where individuals receive instruction, or work together with a more skilled individual who guides them toward learning the information and skills they lack. In the next section, we discuss our model of trait relationships and the learning environment in more detail.

7.2.1 Representation of Traits And Their Prerequisites

In order to represent the "prerequisite" relations between a number of cultural traits, we organize the traits into trees[1], where nodes higher in the tree represent knowledge, skills, or concepts which are necessary for traits further down the tree. Let us consider the different skills and information necessary for the construction of a single artifact, say a dart thrown by an atlatl. An artisan will possess information about different raw materials, an understanding of what materials are suitable for specific purposes, skills and information concerning the knapping of different types of bifaces, methods of hafting bifaces into different kinds of shafts, and so on. Stout (2011) organized such knowledge into "action hierarchies," which represent sequences of actions, sets of choices, and optional elements for the construction of a class of stone tools, drawing the representation from Moore's (2010) graphical notation.

We should emphasize that employing tree structures to represent learning dependencies is a modeling choice. Other choices may be sensible as well. General graphs could represent webs of relations between concepts or skills, and multigraphs (replacing adjacency matrices with tensors) can represent different types of relations in a single structure (Nickel et al., 2011). For purposes of the present chapter, we are interested in the order in which people usually *learn* skills and information, rather than the order in which steps are executed. The difference is potentially significant, in that two adjacent steps in a sequence might involve very different information, tools, or skills, which can

---

[1] A tree is a graph with no cycles or loops. That is, a tree is a connected graph on n vertices that possesses at most $n - 1$ edges (Diestel, 2010). Furthermore, in this chapter we are concerned with *rooted* trees, in which one vertex is distinguished as the "origin" of the tree, giving rise to a hierarchical structure.

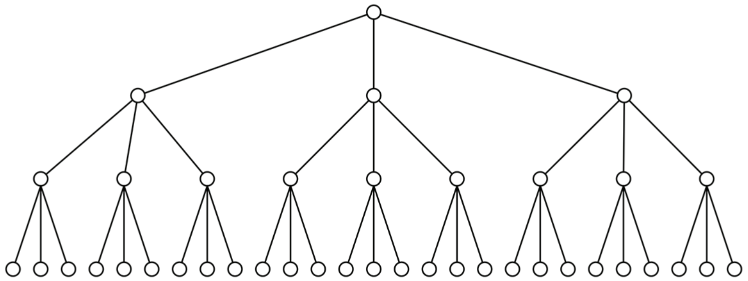


Figure 7.1: A single trait tree, represented by a balanced tree with branching factor 3 and depth factor 3, order 40. In our model, nodes higher in the tree represent prerequisites for nodes lower down the tree. Each instance of the model will have several or many of these trees in the design space.

be learned in parallel without dependencies. Because, in our model, traits cannot be learned unless an individual possesses the necessary prerequisites, we introduce the idea of a "learning hierarchy," which is a division of Stout's action hierarchy into components which are learned with ordered dependencies, and independent components represented in separate trees. For example, one might learn about the sources of good lithic raw materials, independent of learning how to perform different percussion techniques. In our model, each of these independent areas is represented by a separate tree of traits.

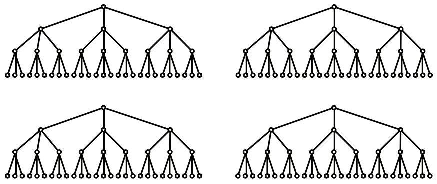


Figure 7.2: A design space composed of 4 independent trees, each tree with branching factor 3 and depth factor 3, order 40. We also studied larger design spaces with 16 independent trees, and with larger branching and depth factors.

In each simulation model, we begin with a trait or "design space" that incorporates several independent sets of traits (O'Brien et al., 2010). The overall design space of a simulation model is thus a forest[2], composed of several trees (Figure 7.2). For each tree in a learning hierarchy, we employ balanced trees which have the same number of nodes at each level, to provide a simplified model of a design space with which to begin our exploration of this class of social learning model, although real design spaces are undoubtedly more complex in their geometry. Each tree in our model is specified by a branching factor r and depth h. As a result, each trait tree in the design space has $\sum_{i=0}^{h} r^i$ traits.

The tree depicted in Figure 7.1 thus has 40 vertices, for example. In this chapter, we examine both small (4 trees) and larger (16 trees) design spaces, to see how learning may differ in problems involving design spaces of different size and complexity. We examine trees with combinations of branching and depth factors of 3 and 5. Thus, a design space with 4 trees with branching and depth factors of 3 (as in Figure 7.1) would have 160 traits, whereas a design space with 16 trees of branching and depth factors of 5 would have a total of 62,496 traits.[3] Even the small design spaces we consider here create a large space for cultural change and differentiation, given the number of possible trees one can construct on even 40 vertices.[4] In the experiments reported here, the overall size of the design space remains constant over time, which is a simplifying assumption as we develop this class of structured information models. In future work, we will explore the role of invention in episodically creating large new regions of design space for the evolving population to explore.

Given the total "design space" represented by a forest of trait trees, each individual in our model is initialized with a small number of "initial" traits. Initial traits are chosen randomly but heavily weighted towards the roots of the trees to represent the fact that our knowledge starts out basic and

---

[2] A forest is a graph composed of multiple components, each of which is a tree.

[3] We initially chose 6 as the limit on branching and depth factors, but found that we cannot calculate certain symmetry statistics, such as the size of the automorphism group, on trees that large using existing tools. Even a tree with $r = 5, h = 6$ has over 10^{1623} possible symmetries, and an attempt to calculate the symmetries for $r = 6, h = 6$ did not complete given the memory limits of the computers we had available.

[4] If we consider each trait to be unique and non-interchangeable, the number of unique trees with unique vertex labels is n^{n-2} by Cayley's theorem (Diestel, 2010). For example, for each trait tree of 40 vertices, there are roughly 10^{60} possible trees. Even if we consider traits to be interchangeable (e.g., we look at the abstract topology of trees rather than the details of individual traits), there are *at least* 10^{16} possible unlabelled rooted trees on 40 vertices (using Otter's (1948) approximation).

sparse. In general, all of the design spaces modeled here are larger than populations will explore within the bounds of a simulation run. In the next sections we describe the social learning model, modified from Robert Axelrod's original, by which each simulated population evolves within this tree-structured design space, and will return to the specifics of how an initial culture repertoire is chosen.

7.2.2 The Axelrod Model of Social Learning and Differentiation

Robert Axelrod (1997) formulated a model aimed at studying the conditions under which simple learning rules could lead to cultural differentiation, rather than a single fixed state (which is the result of simpler neutral or diffusion models). This makes it useful as a starting point for understanding phenomena such as behavioral modernity, in our view. Axelrod's model combines social learning, in the form of random copying, a spatial structure to interaction, in the form of localized copying of neighbors on a lattice, and the tendency to interact most strongly with those to whom we are already culturally similar (homophily). The model displays a rich and interesting set of behaviors, and has been extensively studied by social scientists and physicists (Castellano et al., 2009). First we review the basic model, and in the following section our modified algorithm.

7.2.2.1 Axelrod's Original Model

The original model locates N individuals on the nodes of a regular lattice or grid, but various network structures have also been studied. Each individual is endowed with F integer variables $(\sigma_1, \ldots, \sigma_F)$, that can each assume q values. In the original model, each variable is a "cultural feature" each of which can assume q "traits." In each step, a randomly chosen individual i and a random neighbor j are selected, and "interact" with probability equal to the overlap between their cultural repertoire.

Overlap, in the basic model, is simply the fraction of features for which i and j possess the same trait value:

$$p(i, j) = \frac{1}{F} \sum_{f=1}^{F} \delta_{\sigma_f(i)\sigma_f(j)} \qquad (7.1)$$

where $\delta_{i,j}$ is Kronecker's delta function, taking the value 1 when its two arguments are equal and 0 otherwise. When individuals interact, the focal individual i takes the trait value of its neighbor for one of the features where the two individuals differ.

Interaction has no effect when two individuals already possess identical cultural repertoires, and there is no probability of interaction if individuals have no traits in common. This eventually causes the model to reach an absorbing state where no further changes are possible. Instances of the model are initialized with a random distribution of traits among individuals, and left to update until the steady state is reached. The evolution of the population leads to two classes of absorbing states: (a) a "monocultural" state in which all individuals share the same set of variables, and (b) a "polycultural" state in which subpopulations exist which share the same set of variables within the group, and are completely different from their neighbors.

Which of the two results is reached, and the statistical character of "polycultural" states when they exist, depends mainly upon the number of traits possible q for each cultural feature. For small values of q, individuals share many traits with their neighbors, interactions are thus frequent, and one domain comprising a single set of traits will grow to become fixed within the entire population. In contrast, when the value of q is high, individuals start out sharing very few traits, with interactions that are correspondingly less frequent. Regions of uniform cultural variation do grow, but as they do, sets of individuals who share no traits at all (and thus do not interaction) grow as well, and often prevent any single regional culture from expanding to fix within the population.

Many variants of the basic Axelrod model have been studied, including the addition of "drift" via the introduction of copying error, situating agents on different types of complex networks, the addition of an external "field" to simulate the effects of mass media, and copying that obeys a "conformist" or majoritarian rule by selecting the most common trait among the neighbor set (Castellano et al., 2000; De Sanctis and Galla, 2009; Flache and Macy, 2006; Gonzalez-Avella et al., 2007b,a; González-Avella et al., 2005, 2006; Klemm et al., 2003a,b, 2005; Lanchier et al., 2010; Lanchier, 2012). In general, modifications of the basic model can reduce the tendency of the model to produce polycultural solutions, or change the time scale or location of the critical point.

7.2.2.2 Semantic Extensions to the Axelrod Model

We begin each simulation with N (100, 225, or 400) agents, arranged on a square grid. A design space is created, with some number of trait trees (4 or 16), with uniform branching factors and depth factors (3 or 5). An example of such a tree is shown in panel A of Figure 7.3. Initial traits (and their prerequisites) are chosen randomly across the configured number of trait trees, as follows. For each individual, we select a random number t between 1 and 4, and repeat the trait selection process t times for that individual. In each selection, we choose a random tree in the design space, and then select a depth in the tree for the trait, given by $d \sim \text{Poisson}(0.5)$. This biases trait selection towards the root of the tree, as one would expect in young or inexperienced individuals. We then walk d steps into the tree, making uniform random selections for the children of each vertex. The path of vertices thus constructed is added to the individual's trait set, giving them an initial trait and its necessary prerequisites. One such initial trait is shown in Panel B of Figure 7.3. Given that individuals begin with a small number of initial traits (between 1 and 4, selected randomly), and their prerequisites, the initial trait endowment of an individual is between 1 and 4h, where h is the maximum depth of the design space (either 3 or 5 in the experiments reported here).[5]

---

[5] At maximum, this yields some individuals who begin the simulation with up to 20 traits. The median number of traits in samples taken after 6-10 million time steps is considerably higher–259 traits per cultural configuration or

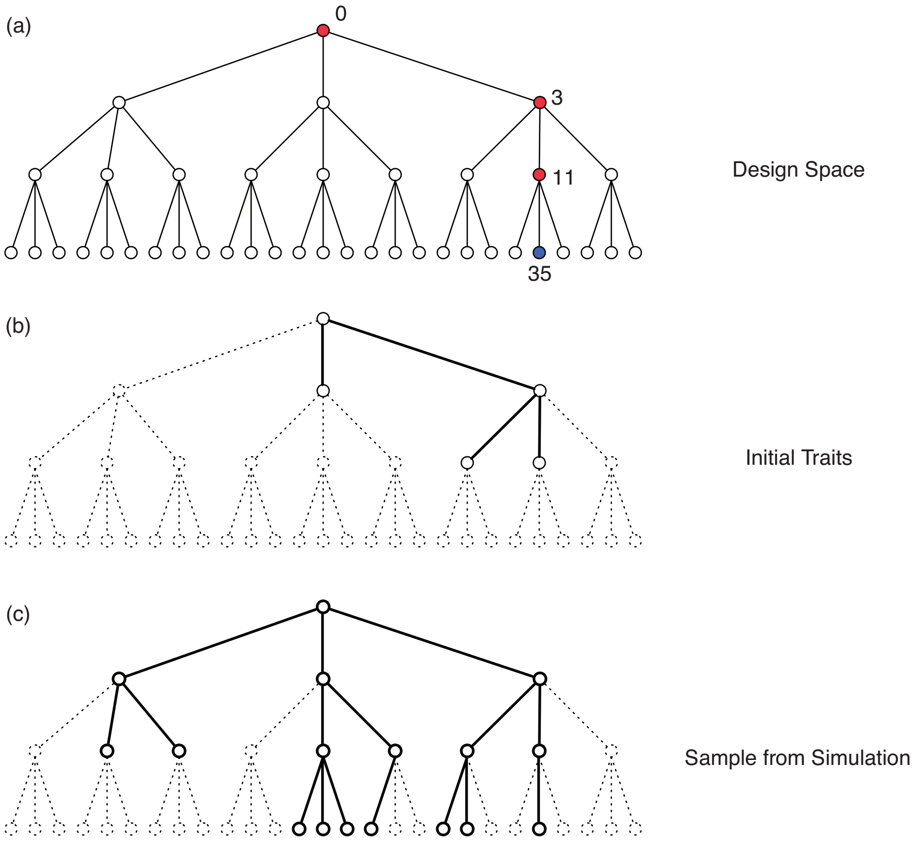


Figure 7.3: Illustration of a design space composed of a single trait tree, along with a random initial trait chosen from the design space, and a final sample from a simulation run, showing the evolution of traits within the design space. Also shown in the top panel are the "prerequisites" for a cultural trait (35), as an example.

Once the population is initialized, the simulation runs a discrete approximation to a continuous-time model. In other words, only one agent changes at each elemental time step, as in the original Axelrod model and the Moran model of population genetics and its cultural version (Aoki et al., 2011; Moran, 1962, 1958). At each step, an agent (A) is chosen at random, and a random neighbor of A is then selected (agent B). Their probability of interaction is given by the overlap of trait sets, which is most simply calculated as the Jaccard overlap between the set of tree vertices each possesses,

---

region. Thus, cultural repertoires in the simulation grow through copying and innovation, as expected.

thus replacing Equation 7.1 with:

$$J(A, B) = \frac{|V(A) \cap V(B)|}{|V(A) \cup V(B)|} \tag{7.2}$$

where $V(i)$ represents the vertex list for trait trees held by individual i in the population.

If the agents end up interacting, agent A observes the traits currently possessed by B, and selects a trait (T) that it does not already possess to learn. If agent A has the necessary prerequisite traits for the selected trait, it can learn trait T. If not, there is a probability $\mathbb{P}(l)$ that B can teach A a necessary prerequisite for T instead. This simulates the process of agent B structuring the learning environment of A through formal instruction or apprenticeship, for example. If such a prerequisite learning event occurs given $\mathbb{P}(l)$, agent A learns the most fundamental of T's prerequisites that it does not already possess. For example, agent A might require the trait closest to T (e.g., trait 11 in Figure 7.3, if the original trait targeted was 35).

Additionally, at each time step, there is a probability $\mathbb{P}(m)$ that one random individual in the population will learn a new trait (and necessary prerequisites) that it does not already possess. For example, if an innovation event occurs and an agent discovers trait 35 by individual trial and error learning, we assume that the agent also discovered traits 0, 3, and 11. Thus innovation can introduce one trait to the population, or a linked set depending upon its prerequisites and what the innovating individual already "knows." This model of innovation simulates an ongoing process of individual learning unconnected to social learning or teaching within the population. Because this functions much like "infinite-alleles mutation" in the classical Wright-Fisher neutral models (Ewens, 2004), or like noise terms in Axelrod, Ising, or Potts models (Castellano et al., 2009), we will refer to this as the "global innovation rate" in this chapter.

One of the editors noted that this model of innovation may not be as realistic as an alternative, where random innovations would be "discoverable" only with the correct prerequisites in place. We believe that innovation in the face of skill or knowledge prerequisites is continuous between these two models. Occasionally one will discover a new piece of knowledge or develop a skill, having learned surrounding and related knowledge. In other situations, individuals may learn sequences and sets of information or skills by trial and error and "tinkering." The "size" of innovations that can be learned purely by individual trial and error should thus vary between these extremes, biased towards the "small" end of the range. Our selection of an innovation model where individuals discover a trait and its prerequisites thus potentially overestimates the effect of individual learning, but it made certain graph operations easier, and can be relaxed in future models.

Each simulation run lasts 10^7 steps, which yields between 10^4 and 10^5 copying events per individual, depending upon population size.[6] Since we do not explicitly model the interaction between cultural transmission and biological reproduction here, we can interpret the model as representing either fine-grained learning within individuals over the course of their lifetimes, or long-term cultural evolution within a fixed size population where we are not modeling fitness. We felt this simplification was appropriate in a pilot study exploring structured information models, but a more detailed study would include dynamics on two time scales: developmental learning and evolutionary dynamics given birth and death. Samples are taken beginning at 6 million steps, and sampling at an interval of 1 million steps, and record the trait trees seen in the population. An example of such a sampled tree is shown in Panel C of Figure 7.3. For reference, the full algorithm for each copying step is given in the Appendix as Algorithm 7.1.

---

[6] 100,000 was chosen as a compromise for running large batches of simulations in parallel. Some simulation runs, especially in small design spaces with very high prerequisite learning rates, can converge to a monocultural solution and quasi-stable equilibrium quite quickly; in the largest design spaces and low learning rates, convergence may never occur even though the process is well-mixed. However, the processes have reached a quasi-stable equilibrium, verified by examining samples at different times for secular trends in median and mean values, which were not found.

7.3 Measuring Cultural Diversity and the Results of Structured Learning

Each sample from a simulation run is composed of the distinct sets of trait trees possessed by individuals in the population, along with summary statistics. If a simulation run converges to a monocultural solution, the sample will have one set of trait trees, which are shared across the entire population. In other cases, there will be clusters of cultural configurations which might be unique to a single individual, or shared by some number of agents. Each cluster will be composed of some number of trait trees (typically, the number configured for the simulation run: 4 or 16, but perhaps a subset), and each trait tree will be the result of many agents learning traits and their prerequisites socially, and for runs with a non-zero mutation rate, by individual learning or innovation. Each cluster will thus have some number of traits, typically higher (often much higher) than the initial endowment given to the population.

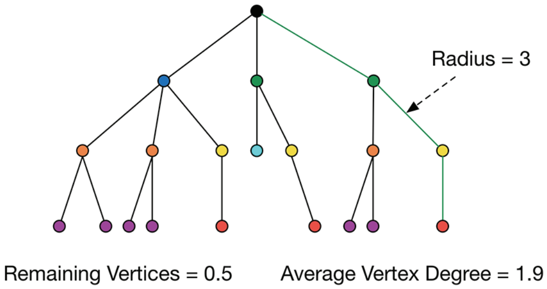


Figure 7.4: An example set of traits at the conclusion of a simulation run, extracted from a simulation with branching factor 3 and depth factor 3, and a single trait tree as the trait space. The remaining density of vertices, mean vertex degree, and radius of the tree are noted. Vertex colors denote "structural equivalence" classes or "orbit structure," as measured by adjacency patterns, and is one measure of the symmetries present in the tree.

From the sampled trait trees, we calculate summary statistics as follows. The ratio of the number of traits in the sample to the full design space size (or "remaining density" of traits) is one measure of trait richness. The radius of a rooted tree is the number of edges in the path from root to the furthest edge. The average radius of trees in a sample (or its ratio to the depth of the design space) is another richness measure, aimed at measuring whether knowledge with multiple prerequisites is being learned within the simulated population. Similarly, in the original design space, the branching factor describes how many children each node in the tree started with, so measuring the average vertex degree gives us a rough measure of how broad a cultural repertoire is. Each of these measures is illustrated in Figure 7.4 for an example tree selected from our data.

In addition to these simple numerical measures comparing final trees to the original design space, it is useful to measure something about the overall "shape" of the trees themselves. One way of formalizing this notion is to examine the *symmetries* of the final trait trees.

Examining Figure 7.4, if we ignore the exact identities of traits for the moment, it is apparent that there are repeating patterns. For example, the left-most branches each terminate in a pair of leaves. This pattern is repated on the second right-most branch. These types of repeating patterns are computationally expensive to search for in large sets of trees, but we can summarize them by considering trait trees as algebraic objects and examining their *automorphisms*.

Figure 7.5 provides a simple example of symmetries in a small tree, and how those symmetries are "lost" as the tree is thinned of vertices and edges. Thus, measuring the total number of interchangeable patterns or symmetries in a tree representing a cultural repertoire is a relative measure of how a repertoire "fills" the original design space. An automorphism is a function which maps an object to itself, in such a way that the structure of the object is preserved (Rotman, 1995), and thus is an operational definition for the "symmetries" one intuitively perceives in something like the trees shown here. Graph automorphisms map vertices in a graph to each other, preserving properties such as the adjacency pattern of edges.

For purposes of relative comparison, we measure the total number of automorphisms found in each sampled cultural repertoire. In algebraic graph theory, this is referred to as the size of the auto-

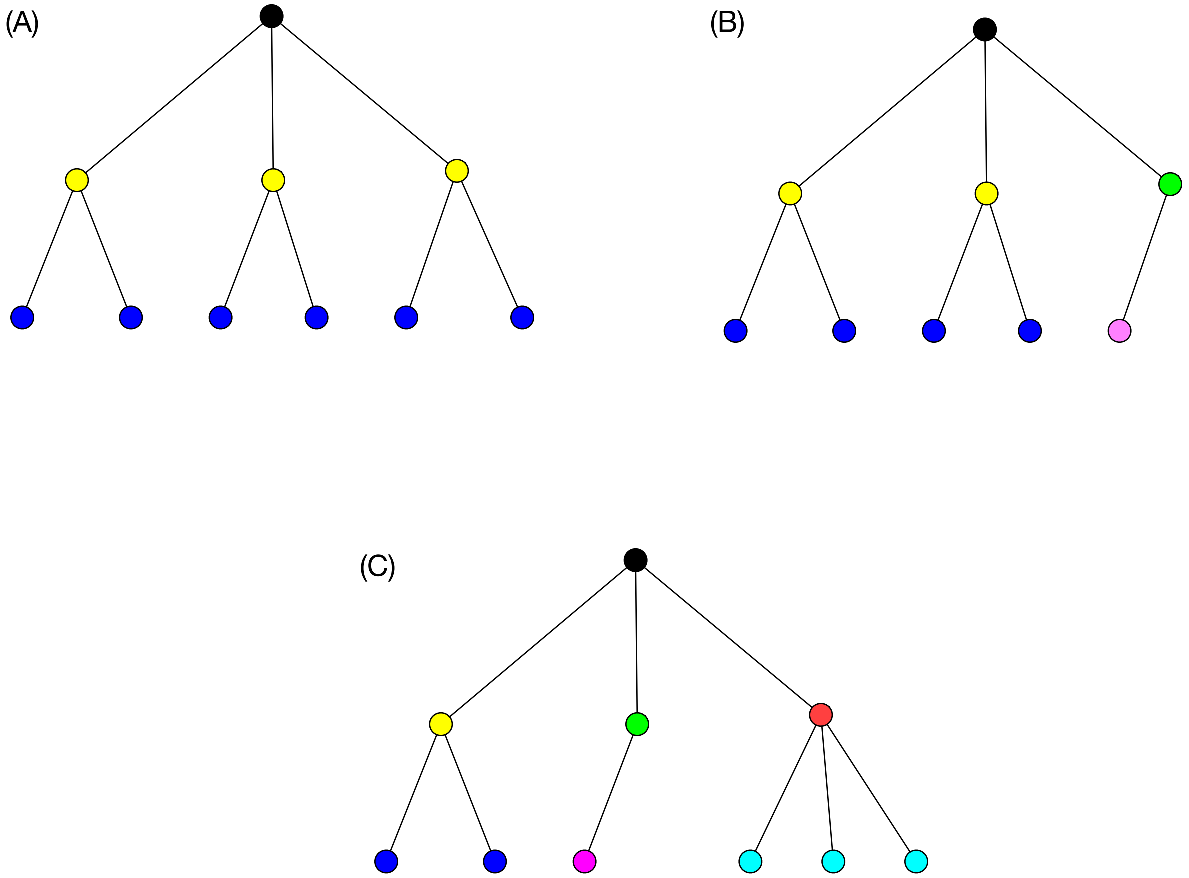


Figure 7.5: An illustration of symmetries present in a graph. Vertices that occupy positions with the same connectivity structure are interchangeable without changing the connectivity structure, and thus represent a "symmetry" under a transformation, while leaving the overall structure of the graph unchanged. The full tree (A) displays the most symmetry, with all; vertices at each vertical level "swappable for each other". As vertices are deleted in graphs (B) and then (C), vertices become less and less interchangeable given different numbers of edges present, and thus are broken into non-interchangeable sets, shown as different colored vertex sets (technically, each set is an "orbit" of the automorphism group of the graph).

morphism group or $|Aut(G)|$ (Godsil and Royle, 2001). A tree with no repeating patterns will thus have an automorphism group size of 1, indicating that the only symmetry is the entire tree itself. A balanced tree with branching and depth factors of 3, as depicted in Figure 7.1, has approximately 1.3×10^{10} automorphisms. The more repeating patterns there are in trait trees, the more automorphisms they will possess.

Because group sizes grow quickly and the accuracy of performing calculations with truly astronomical numbers is low, another possible measure of the symmetries present is to count the *classes* of equivalences into which vertices fall. The *orbits* of the automorphism group are the sets of vertices which are interchangable by some permutation that preserves structure. For example, the graph in Figure 7.1 has five orbits, with each vertex at a given level interchangable (in a structural sense). Similarly, the six leaf vertices that are part of pairs in Figure 7.4 are part of the same orbit; in this illustration, each orbit is given a different color to highlight their equivalence. For each cultural region found when sampling a simulation, we calculate the size of the automorphism group and the number and multiplicity (frequency) of orbits. For this analysis, we employ the *nauty + Traces* software by Brendan McKay and Adolfo Piperno (McKay and Piperno, 2014).[7]

7.4 Experiments

Given a modified Axelrod model on a tree-structured trait space, we expect to see greater cultural diversity, differentiation among groups of individuals, and larger sets of traits as the "learning environment" is tuned from a low to high probability of teaching and learning among individuals. We also expect that individual innovation, independent of the social learning context, will increase the amount of the technological design space that a population explores, which leads to enhanced

---

[7] Nauty+Traces can be downloaded at `http://pallini.di.uniroma1.it/`. We employed version 2.5r7 for this research.

opportunities for differentiation even through simple random copying. Here we measure cultural differentiation by the number of clusters of individuals who share the same trait trees when we sample the population.

Second, we looked at whether highly structured learning environments, represented here by higher probabilities of naive individuals gaining the prerequisites for the skills and information they encounter with peers, led to deeper and richer cultural repertoires. We explore a number of ways of measure the richness of a cultural repertoire in a model with structured relations between traits, through the use of graph properties and symmetry measures. The measures used are those described above: the tree radius (or depth), mean vertex degree, the fraction of remaining vertices, and the size of the automorphism group of sampled trait forests. Finally, we began to examine how the structured learning environment might interact with demography, by simulating the same parameters across two sizes of population.

For this chapter, we examined populations of size 100, 225 and 400, to begin to examine the effects of population size. For these populations, we examined design spaces that were small (4 trait trees) and large (16 trait trees). Within each size, we further examined combinations of branching factor and depth factor with values of 3 and 5, thus yielding 8 total sizes of design space (Table 7.1).

Branching Factor	**Depth Factor**	**Number of Trait Trees**	**Size of Design Space**
3	3	4	160
5	3	4	624
3	5	4	1456
5	5	4	15624
3	3	16	640
5	3	16	2496
3	5	16	5824
5	5	16	62496

Table 7.1: Size of design space for different trait tree configurations

Further, we examined three levels of global mutation or innovation rate: zero, or no mutation, and 0.00005 and 0.0001. Such rates created a constant supply of new innovations, but several orders of magnitude less frequent than copying and prerequisite learning events. The full set of parameters are given in Table 7.2. In this pilot study, for each combination of all of the above parameters, we performed 25 replications. With 5 samples per simulation run, this yielded 10,963,691 samples of cultural regions.

Simulation Parameter	**Value or Values**
Population rate at which new traits arise by individual learning	0.0, 5e-05, 0.0001
Maximum number of initial traits (not including their prerequisites) each individual is endowed with	4
Number of distinct trees of traits and prerequisites	4, 16
Population sizes	100, 225, 400
Replicate simulation runs at each parameter combination	25
Maximum time after which a simulation is sampled and terminated	10000000
Individual probability for being taught a missing prerequisite	0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
Number of branches at each level of a trait tree	3, 5
Depth of traits in each trait tree	3, 5

Table 7.2: Parameter space for simulations described in this chapter

7.5 Results

We begin by noting that compared to the original Axelrod model, or neutral and biased copying models, the dynamics of our semantic Axelrod model are highly variable. A very wide range of outcomes is possible for each parameter combination, especially when the size of the design space is large. Some variables, such as the average vertex degree of sampled trait trees, are strongly overlapping across all learning rates and do not appear diagnostic of different learning environments, at least in these initial experiments. Given the large amount of variability in the dynamics, larger numbers of replications would be useful, although this is computationally quite expensive at present.[8] That said, several features of the data are strongly suggestive that hierarchical trait models have potential in modeling cumulative technological evolution, making the computational expense worthwhile.

7.5.1 Cultural Diversity

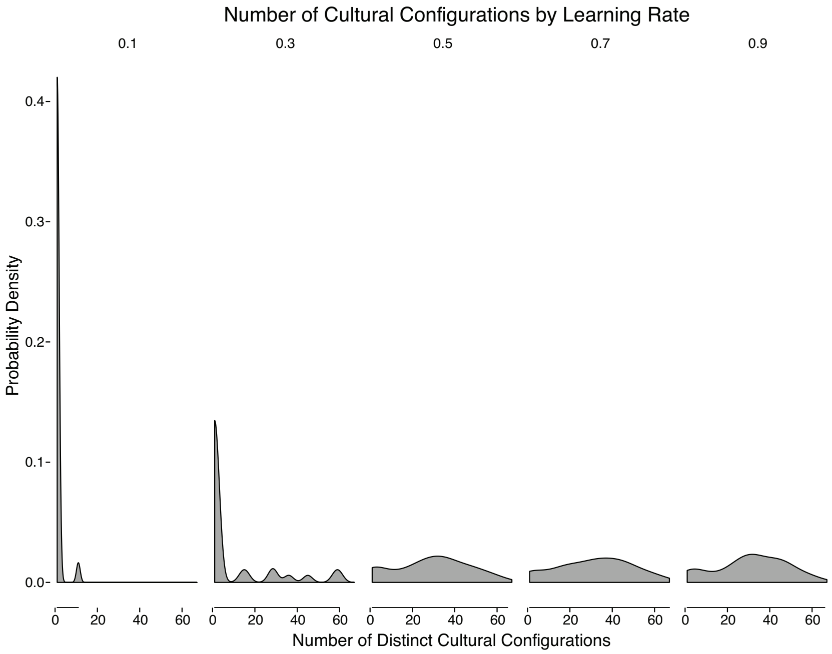


Figure 7.6: Number of cultural configurations in simulations with the smallest trait space (160 total traits in 4 trees), and a high individual innovation rate (10^{-4}).

Variation among individuals is foundational to evolutionary processes, and is the raw material from which differentiation between regions and cultural groups is constructed. Figure 7.6 depicts the number of cultural configurations (i.e., trait trees) in a population of size 100, for the smallest

---

[8] The simulations reported here ran on a cluster of 6 compute-optimized "extra large" Linux instances on Amazon's EC2 computing cloud, for a total of 17 days of wall clock time and 2075 CPU hours. We plan further optimizations to the simulation code to make larger samples economically feasible.

trait space with only 160 total traits, and relatively high levels of individual innovation. For example, in the left-most panel the large peak just above zero indicates that most simulated populations are characterized by one or a few sets of trait trees. Five learning rates are depicted, increasing from left to right across the panels. At the very lowest rate of learning fidelity, with only a 10% chance of being taught a needed prerequisite for knowledge being copied, most of the populations simulated share a single set of traits, and even individual innovation does not drive significant exploration of the space of structured traits. With increased fidelity in teaching needed prerequisites, however, simulated populations begin exhibiting marked differentiation, with individuals possessing more unique configurations of traits from the overall design space.

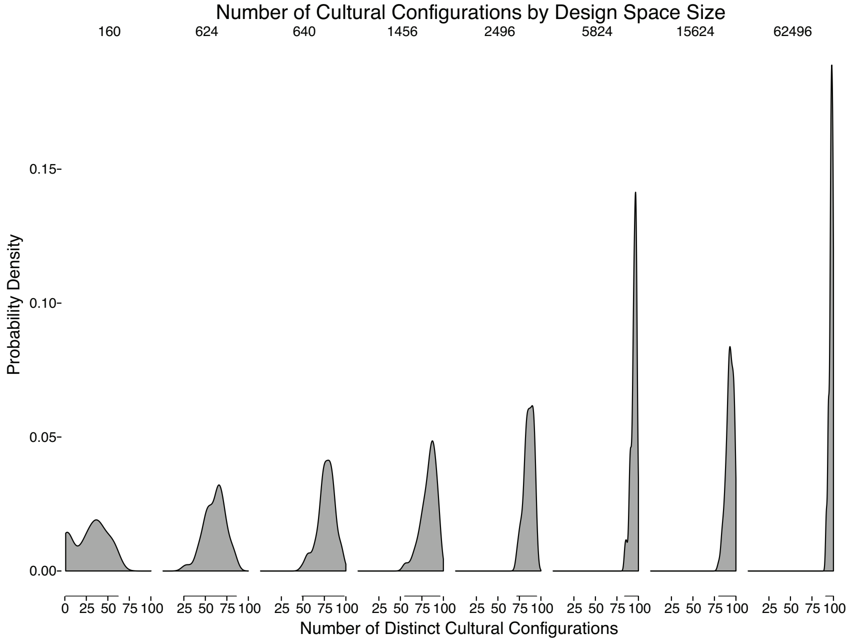


Figure 7.7: Number of cultural configurations in simulations with an intermediate learning rate (0.4), across different sizes of trait space.

Looking at the data from another perspective, we can hold the fidelity of learning constant (say, at a 40% chance of being taught a needed prerequisite), with the same global innovation rate (10^{-4}) as Figure 7.6, and examine the effect of different size design spaces (Figure 7.7). In general, populations exhibit greater differentiation between individuals as the design space gets larger, as prerequisite learning helps individuals acquire adjacent traits, and individual innovation randomly explores more distant portions of the design space.

Given the structure of the Axelrod model, with the strong tendency towards cultural uniformity given homophily, all simulated populations converged to a single cultural configuration in the absence of a global innovation rate. This highlights the importance of various "innovation" and "invention" processes in the creation and maintenance of cultural differentiation and diversity (Eerkens and Lipo, 2005; O'Brien and Shennan, 2010), and suggest that highly conservative cultural repertoires, such as those posited to precede behavioral modernity in hominin populations, occur whenever individuals engage in social learning in small technological design spaces, in the absence of strong and regular individual innovation.

7.5.2 Trait Richness and Knowledge Depth

Cumulative evolution of technology is represented in our model by the population learning its way *down* the trees which compose the design space. Possession of traits deeper in the trees represents skills or information which is more specific, possessing more prerequisites. Thus, we expect that the depth (or "radius", see Figure 7.4) of trees would increase with the prerequisite learning rate, representing a learning environment which is structured to ensure such acquisition.

Figure 7.8 gives the *normalized* mean radius of cultural regions, broken out by the prerequisite learning rate along the horizontal axis, and each group of 3 boxplots displays the differing global innovation rates studied. Radii are normalized to the depth of their design space, to facilitate comparison. The results indicate that essentially two regimes exist: shorter trees, which do not grow much beyond their initialized size, and larger trees. The mean radius has an asymptote just above 0.75, achieved with the prerequisite learning rate is approximately 0.4 or higher. Further increases

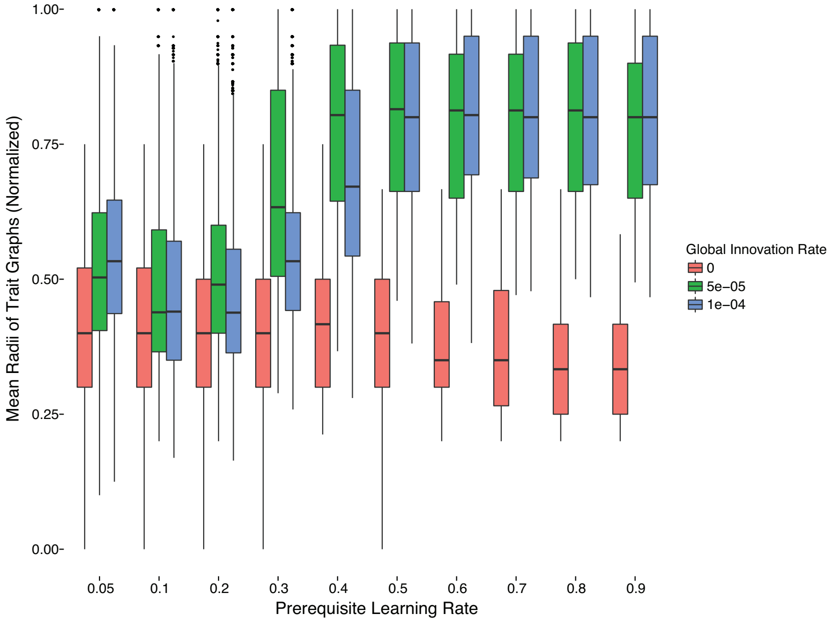


Figure 7.8: Mean depth of trait sets, by prerequisite learning rate and global innovation rate, for population size 100.

do not seem to matter. Additionally, the difference between the two global innovation rates is small–what matters most in terms of qualitative behavior is the presence of global innovation outside the teaching or learning of prerequisites themselves.

7.5.3 Population Size

Earlier, we mentioned that population size does not seem to be a primary factor in explaining the measured diversity in cultural transmission models, except perhaps in bottleneck situations like the one Henrich analyzes in Tasmania (2004). Instead, population size may have an interaction effect with other factors, yielding smaller second-order effects. We examined the effect of population size in the research reported here, repeating the entire set of simulation runs for populations of 100, 225, and 400.[9]

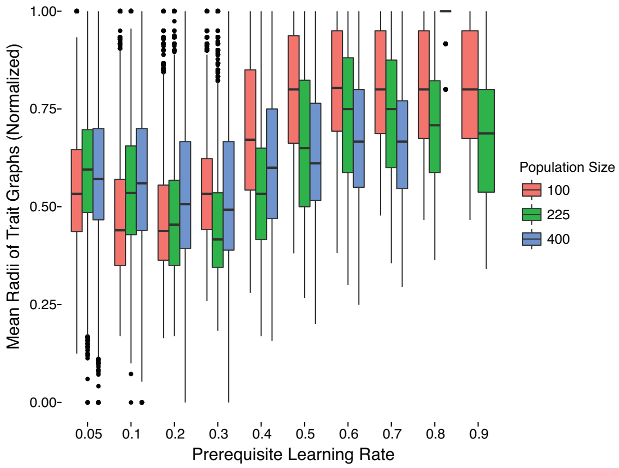


Figure 7.9: Mean depth of trait sets, by prerequisite learning rate and population sizes of 100, 225 and 400.

Figure 7.9 displays the relationship between mean radius (or depth) of the cultural traits in each cultural sample, as in Figure 7.8 above, but the boxplots are instead colored by population size. At least over a range of group or deme sizes likely to be relevant to Paleolithic archaeology, population size makes no difference to the qualitative behavior of the model. There is, however, a very slight decrease in mean radius of trait sets with larger population size, which is likely a consequence of a larger population spreading out over the trait space.

7.5.4 Trait Tree Symmetries

Finally, we examined the algebraic properties of the trait trees composing cultural regions, examining both the number of vertex equivalence classes (orbits) and the size of the automorphism group of the trait forests. We examined the raw metrics, and versions normalized by the size of the maxi-

---

[9] We should note that learning rates of 0.8 and 0.9 for population size 400 were cut short due to budget constraints, but this does not appear to affect the pattern in our dataset.

mally symmetric forest with the same number of traits, branching factor, and depth factor. The latter proved difficult and led to serious overflow problems even with 64 bit arithmetic, so we focus here on the raw automorphism group size.

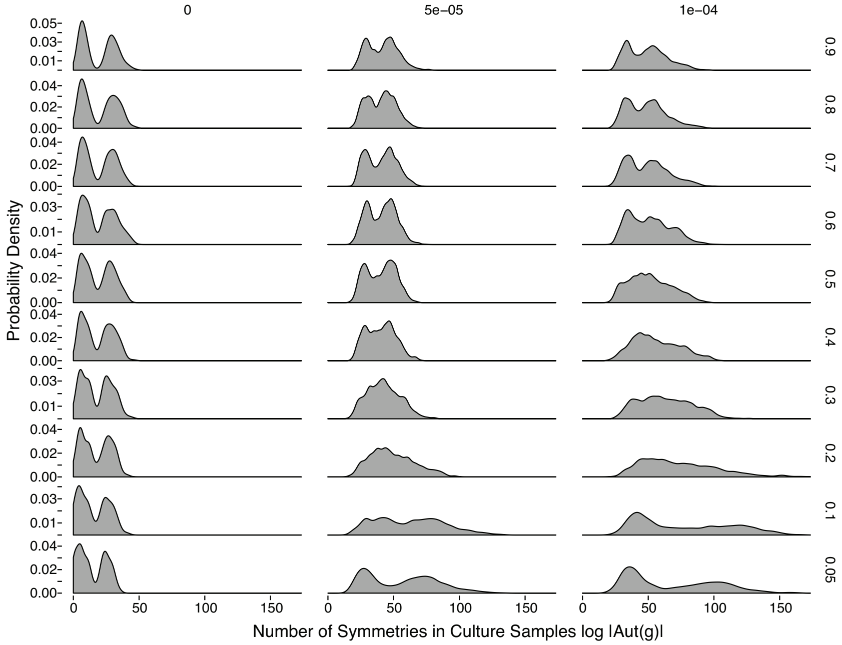


Figure 7.10: Number of symmetries in trait tree samples, measured as the log of the order of the automorphism group of the trait graphs, broken down by prerequisite learning rate (rows) and global innovation rate (columns).

The logarithm of the automorphism group size does hint at interesting structure (Figure 7.10). In the presence of mutation, the learning of prerequisites narrows the range of variability for the automorphism group size, and at higher learning rates renders the distribution multimodal. The modality arises because of the different combinations of branching factor and depth factor we employed for design spaces–i.e., some design spaces are "wide" and some are "narrow," while also being "shallow" or "deep." This gives rise to different modes in the measured symmetries, but overall the reduction in variability in symmetry is the most important qualitative effect seen in our data.

We do not fully understand the "shapes" of cultural regions to which the model appears to converge, but it appears that there is a tendency for trait graphs to converge towards shapes which have moderate numbers of symmetries. This graph is on a logarithmic scale, so a peak at 50 along the horizontal axis correponds to a trait graph with approximately 5×10^{21} symmetries. This is a fairly small number, compared to the original design spaces, which have symmetries ranging from approximately 10^{41} to 10^{6496}. Thus, the geometry of cultural traits in our hierarchical design spaces are fairly asymmetric and represent small and very specific segments of the total design space.

Further analysis of trait graph "shapes" is needed to tell whether there are repeating patterns or graph "motifs" which characterize a social learning model in a graph-structured trait space. The results here are suggestive of such a phenomenon, but inconclusive given just the bulk algebraic properties of cultural regions, since the size of the automorphism group (or the number of orbits) tells only *how many* symmetries there are, not what types of symmetries exist. The next step in our analysis of shape is to pursue a geometric decomposition of the graph following Ben MacArthur and Rubén Sánchez-García's (2008) work on the symmetries of complex networks.

7.6 Discussion

The "semantic Axelrod" model described here specifically addresses social learning of knowledge with "prerequisite" structure, and a learning environment which is tunable from low to high fidelity, simulating the intensity with which "teaching" occurs in addition to imitative copying. The model displays a characteristic increase in the cultural repertoires of individuals, as they learn in environments of higher fidelity. At the individual level, an increase in higher fidelity learning within structured information environments both creates path-dependency in what is learned, and increases the chances for specialization among individuals. Hominin populations in which complex knowledge is taught systematically along with prerequisites will accumulate and retain skills and technology faster and to a greater extent than those groups which rely upon natural pedagogy and imitation for social learning.

Previous research had established the importance of teaching and learning environments for cumulative cultural evolution and cultural diversity (Aoki, 2013; Castro and Toro, 2014; Creanza et al., 2013; Nakahashi, 2013). Our contribution in this paper is a model capable of connecting the fact of teaching with the actual structure and content of cultural knowledge. Such models, we believe, are important in explaining the explosion of cumulative material culture that accompanies behavioral modernity. The model described here only makes a start on modeling the additive and recombinative complexity of real technologies, but it does display accumulated depth of "knowledge" or "skills," as represented by the radius or depth of trait trees. In combination with realistic models of technology–such as the production sequences studied by experts on stone tools–we believe that empirically sufficient models of the evolution of specific technologies are possible and within reach.

Several areas suggest themselves for future research in structured information or "semantic" cultural transmission models. Some we are pursuing, others remain open questions and we invite collaboration towards their solution.

- Regional scale cultural differentiation given a metapopulation embedding of the basic model.
- Additional trait relations (e.g., class subsumption, functional equivalencies).
- Realistic technology models for key artifact classes (e.g., bifaces, scrapers, pottery).
- Incorporation of trait fitness in order to study directional change.

Models of the class introduced here are "thicker" descriptions of how humans acquire skills and information in real learning environments, and thus complement existing models which describe the conditions under which teaching and structured learning might evolve and spread. We believe models of this type make a needed "downpayment" on cultural transmission models which can substantively incorporate specialties such as archaeometry, the technnological analysis of lithics and pottery (Tostevin, 2012), and studies of how innovation occurs in various tool classes (e.g., O'Brien and Shennan, 2010). Bringing cultural transmission modeling together with the details of technologies will be a crucial component in multifactor evolutionary explanations for the complex of changes seen in modern *Homo sapiens* and some Neanderthal populations in the later Paleolithic.

7.7 Acknowledgements

The authors wish to thank Briggs Buchanan and Mark Collard for the invitation to participate in the symposium "Current Research in Evolutionary Archaeology," at the 79th Annual Meeting of the Society for American Archaeology in Austin, TX. A summary of this research was presented in that session, and Alex Mesoudi provided valuable comments on an early post-conference draft. Kenichi Aoki and an anonymous reviewer provided feedback prior to publication, and although we did not take all of their suggestions, the comments led to a number of improvements. Madsen wishes to thank Frédéric Chapoton of the Institut Camille Jordan for answering a question about the maximal automorphism group of trees.

7.8 Appendices

7.8.1 Algorithm Description

Algorithm 7.1 describes the “semantic” Axelrod model variant studied in this chapter. Within the algorithm, there are several functions which find traits with particular properties. Some, like **GetTraitUniquetoFocal**(), are fairly simple set operations but were abbreviated to clarify the notation.

Algorithm 7.1

```
Require: innovrate is the population rate at which individuals randomly learn a trait
Require: learningrate is the probability of learning a missing prerequisite during a learning interaction
focal $\leftarrow$ GetRandomAgent()
neighbor $\leftarrow$ GetRandomNeighbor(focal)
if focal $=$ neighbor $\vee$ focal $\cap$ neighbor $= \varnothing$ $\vee$ neighbor $\subsetneq$ focal then
    exit { No interaction is possible, move on to next agent }
end if
prob $\leftarrow$ $($focal $\cup$ neighbor $-$ focal $\cap$ neighbor$)/$focal $\cup$ neighbor
if RandomUniform() $<$ prob then
    differing $\leftarrow$ neighbor $\setminus$ focal
    newtrait $\leftarrow$ GetRandomChoice(differing)
    if hasPrerequisiteForTrait(focal, newtrait) = True then
        replace $\leftarrow$ GetTraitUniquetoFocal(focal,neighbor)
        focal $\leftarrow$ focal $\setminus$ replace
        focal $\leftarrow$ focal $\cup$ newtrait
    else
        if RandomUniform() $<$ learningrate then
            prereq $\leftarrow$ GetDeepestMissingPrerequisite(newtrait, focal)
            focal $\leftarrow$ focal $\cup$ prereq
        end if
    end if
end if
if RandomUniform() $<$ innovrate then
    focal3 $\leftarrow$ GetRandomAgent()
    innovation $\leftarrow$ GetRandomTraitNotInFocal(focal3)
    focal3 $\leftarrow$ focal3 $\cup$ innovation
end if
```

GetDeepestMissingPrerequisite() is a procedure which takes the trait set of an individual, and a trait for which the individual is known to be missing necessary prerequisites, and returns the "most basic" missing prerequisite for that trait (i.e., closest to the root). This is done by finding the path which connects the root and desired trait, and walking its vertices from the root downward, checking to see if each vertex is part of the individual's trait set. The first trait not found in the individual's repertoire is returned.

7.8.2 Availability of Software and Analysis Code

The simulation software used in this chapter is available under an open-source license at Mark Madsen's GitHub repository `https://github.com/mmadsen/axelrod-ct`. Required libraries and software are listed in the source archive itself, and include Python 2.7 and the open-source MongoDB database engine to store simulation output.

The codebase consists of a set of library modules which implement the shared and unique aspects of each model, unit tests to verify the basic functionality of the code, and scripts which execute each model. The **axelrod-ct** repository contains three models:

- An implementation of the original Axelrod model using the **axelrod-ct** libraries.
- A basic model with an "extensible" trait space but no relations between traits.
- A "semantic" Axelrod model with tree-structured trait space representing prerequisite relationships between traits.

Stepwise extension from the original Axelrod to the semantic models on the same code library allowed a degree of verification, which is difficult in a situation where there is no existing mathematical theory against which to compare the code implementation (Committee on Mathematical

Foundations of Verification Validation and Uncertainty Quantification, National Research Council, 2012).

The analysis and final dataset reported here are available, along with the source of this paper and associated presentations, in an associated GitHub repository: `https://github.com/mmadsen/madsenlipo2014`. Statistical analyses of the final dataset were performed in R, rendering our results reproducible given simulated data from the "axelrod-ct" software linked above.

CHAPTER 8

Conclusion and Directions for Future Research

8.1 Conclusions and Contributions

Our primary job as evolutionary archaeologists is to document the course of cultural evolution, by mapping heritable continuity from archaeological remains. That job employs many tools, both new and old, from phylogenetic and cladistic methods still under development today, to seriation and classification methods now more than a century old. Cultural transmission, broadly speaking, is the central organizing concept which explains why there is heritable continuity to be found in the spatial and temporal variation in artifact form. It is natural for us to employ mathematical models of diffusion and transmission to attempt to understand cultural transmission better.

In Chapter 1 I described three different research programs within archaeology on cultural transmission, distinguished by the spatiotemporal scale of their focus. The first, a "microevolutionary" program, stemmed from two different sets of questions. The first sought to determine whether particular assemblages or groups of assemblages showed evidence of selection, by testing the null hypothesis that class frequencies met some criterion derived from neutral models borrowed from population genetics. The second sought to go beyond simply testing for neutrality and attempt to fit psychological models of biased social learning, derived from Boyd and Richerson's seminal work, to assemblage data. Although neither question is necessarily synchronic by nature, the models and statistical approaches employed have tended to be. In Chapters 2 and 3 I studied the question of whether the microevolutionary program could be made empirically sufficient and thus useful for answering detailed questions about the archaeological record. In Section 8.1.1 I review my conclusions and highlight contributions from this work which have not appeared elsewhere in the literature.

Much of that work was methodological and critical in nature. It has been all too easy for our evolutionary modeling of cultural transmission to slip into synchronic reconstructionism that does not take advantage of our unique strength: time depth. The mesoscopic and macroevolutionary programs described in Chapter 1 are both broadly diachronic in focus, and differ mainly in spatiotemporal scale and the methods uses. My main contribution in this dissertation has been to further develop elements of a true "mesoscale" approach to mapping heritable continuity, distinct from the strong focus of macroevolutionary analysis on the use of phylogenetic techniques. In Sections 8.1.2 and 8.1.3 I review my conclusions and highlight unique contributions from this research program. Further, in Section 8.3, I return to the relationship between the "mesoscale" and macroevolutionary programs, and discuss the key role of classification in determining where a particular research problem lies in the continuum between "mesoscopic" and macroevolutionary scales.

The "mesoscale" approach I have sought (in collaboration with Carl Lipo) to develop employs time-honored archaeological methods, updated for a new century and the computing power we now have available, but in conceptual terms is still in its infancy. We are still exploring what the right structure is for transmission models at this scale, given different kinds of archaeological situations. There is considerable scope to elaborate not only the methods described in this dissertation, but develop new observable units and models. In Section 8.2, I outline what I believe to be fruitful next steps to advance elements of this research program.

8.1.1 Microevolutionary Cultural Transmission Models in Archaeology

Despite the fact that archaeology understands itself as studying cultural *change* through the archaeological record, we routinely (and often inadvertently) find ourselves describing the past in ways that are synchronic and essentialized. Even though Nels Nelson (1916) clearly articulated a continuous and diachronic view of stylistic change in his work at San Cristóbal, later pratice in "culture history" framed our knowledge of the past as a series of synchronic snapshots. *Change* through time became represented as mere *difference* between period or phases, and theoretical effort shifted to the reconstruction of the "Indian behind the artifact" (Braidwood, 1959, 79).

The pull of synchronic description and the reconstructionist enterprise of describing the state of a population at a moment in time is no accident—it is built into the way we perceive the world and organize our knowledge. Dunnell (1982) described the cognitive biases that lead to this tendency as our "common sense", which constructs views of the world consistent with the short time scales over which we need to perceive the world and adapt to it. Social and cognitive psychologists have since begun to document the proximate causation for essentialized cognition in early childhood development (e.g., Rhodes et al., 2012; Rhodes and Mandalaywala, 2017; Gelman, 2004).

Indeed, our tendency to think about the world in synchronic and essentialized terms is so strong that Mayr, Lewontin, and others have argued that the primary revolution Darwin kicked off was not about evolution itself, or even the theory of natural selection, but through the introduction of "population thinking" as an alternative to the typological, essentialist thinking about species and the nature world that characterized pre-Darwinian biology (Dunnell, 1982; Lewontin, 1974; Mayr, 1959). The "materialist revolution," as Lewontin termed it, treated *variation* as causal, rather than noise, and required one to think about change in fully diachronic terms. Evolutionary archaeology began with Dunnell's articulation of exactly this point (Dunnell, 1978, 1980, 1982, 1989).

Nevertheless, the application of cultural transmission models within evolutionary archaeological very quickly took on a synchronic, "reconstructionist" flavor. By adopting equilibrium models from classical population genetics, whether the Wright-Fisher model of drift or the social psychological models from Boyd and Richerson's justifiably influential work (1985), the very structure of the models themselves gave predictions for the stationary, unchanging state of a population. Fitting such models to observational data requires treating our data "as if" they were a synchronic sample of a population at a moment in time, so that the observed frequency distribution can be compared to the theoretical one. Nearly all of the microevolutionary program for cultural transmission modeling in archaeology has proceeded in this way, until very recently (Kandler and Shennan, 2013).

Chapter 1 describes in more detail how, as researchers began to see conflicting results when reanalyzing the same canonical data sets over and over in their methodological papers, it became apparent that efforts to fit cultural transmission models to data led to ambiguous conclusions. This realization led to serious discussion about equifinality between transmission models (Barrett, 2019; Kandler and Crema, 2019; Premo, 2010). The first set of questions that comprise this dissertation address this question: how severe are the equifinality problems faced by the microevolutionary approach? Is it possible to distinguish between transmission models using coarse-grained data on frequencies of cultural traits in a population?

Archaeological data are "coarse grained" in two ways, both relevant to the microevolutionary program. The first is that archaeological observations are always diachronic in nature and inherently represent counts and frequencies that refer to depositional events over a duration. Sometimes the duration might be short, as in the archaeological record of recent historical periods, or a single assemblage may represent deposition across millennia. Second, as archaeologists we nearly always operate with assemblages which represent whole populations or subpopulations, except in specific depositional and many historical contexts. This means that heterogeneity within populations may not be distinguishable given assemblage-level data. Furthermore, this means that realistic models of heterogeneous social learning may not be identifiable and empirically sufficient with population level data alone. Chapters 2 and 3 addressed each of these types of "coarse graining" on our ability to fit cultural transmission models. I summarize their conclusions in turn.

8.1.1.1 The Centrality of Time Averaging To Evolutionary Modeling

Chapter 2 examined the effect of time averaged observations on the observable statistics we use to measure goodness of fit between archaeological frequency distributions, and the theoretical models we employ as hypotheses. Since most of our hypothesis tests or goodness of fit testing is related to the shape of frequency distributions, the effect of aggregation on diversity measures is critical to understand. My results indicate that richness is inflated in samples with longer duration compared to synchronic observations. This is the underlying reason why culture historians noted that assemblages used in seriation should be of equal duration. But this fact takes on critical importance with most of our cultural transmission models, since the number of expected traits (or alleles, given their origin in population genetics) is the "sufficient statistic" at a given population size and mutation rate. The "evenness" of frequencies is flattened with greater assemblage duration; this causes problems with neutrality tests that seek to employ the shape of the frequency distribution, including Ewens and Watterson's tests and Slatkin's "exact" tests for neutrality. As a result, such tests display increased Type I error rates.

The most important contribution of Chapter 2 is the linkage between how long cultural traits persist in the record, and the timescale over which time averaging effects appear. Assemblages whose duration is longer than the mean lifetime of the classes which analytically represent cultural variation will display increased richness, flattened diversity, and increased Type I error in "neutrality tests." Assemblages whose duration is equal to the mean trait lifetime, or shorter, are not affected by time averaging in these ways measured here. This is an important result because archaeologists can, at the time of data collection and subsequent artifact analysis, exert some control over the effects that time averaging might have on our analyses. Relative comparisons between assemblages are possible, if durations are comparable. Sometimes it is possible to control the degree to which we aggregate artifacts into "assemblages" for analysis both during fieldwork, and afterwards in an analytical manner.

In like manner, we have partial control over the relative lifetimes of the artifact classes we use to measure cultural variation. We must always remember that the classes and types whose frequencies we are comparing to transmission models are analytical constructions, rather than being inherent in the rocks and sherds we handle (Dunnell, 1971). Because we form the classes we then use for counting abundance and forming frequency distributions, we can vary classification "level" (*sensu* Dunnell 1971) and by doing so, often vary the time span over which a class is represented in the record. Varying the number of dimensions of variation in our classification will affect the measured "lifetime" of the classes themselves. Individual attributes (or modes) belonging to a single dimension of variation (e.g., straight rims on ceramic bowls) may persist in the archaeological record for long periods of time (and, indeed, by themselves may occur in unrelated contexts, thus being non-homologous). But when we combine dimensions of variation, say by constructing a ceramic classification by intersecting rim form with surface treatment with rim decoration, each combination of these dimensions tends to have a restricted spatio-temporal distribution. Since we can affect our classifications and the level of detail they differentiate, we have some ability to tune our observations of cultural trait frequencies in response to the level of temporal aggregation present in the archaeological deposits we seek to study.

This does not mean that careful tuning of classification level, and reduction of assemblage duration is sufficient to render synchronic, microevolutionary models empirically sufficient. Even short duration assemblages are still a diachronic source of data, that demand diachronic, historical explanations. The effects of time averaging are worth studying not because we can "correct" for them, but because aggregation effects tell us something about the time scales our diachronic hypotheses must be framed at, in order for our data not to "underdetermine" alternative models (Perreault, 2019). The study of time averaging effects is integral if an evolutionary archaeology is to develop testable explanations for the archaeological record, as it is critical for paleobiology in developing testable explanations of the fossil record (Kowalewski, 1996).

8.1.1.2 Coarse Grained Data Cannot Distinguish Realistic Microevolutionary Models

Chapter 3 examined the second problem that the microevolutionary approach faces given only coarse grained data sources. Data on artifact class frequencies nearly always reflect prevalence in some population aggregate, with only limited information about individual variation (except in historical archaeology and certain other specialized depositional circumstances). Since real human (and animal) populations are a mixture of individuals with different social learning strategies, we need to understand the degree to which different mixtures of social learning rules are reflected in population-level data in ways that can be statistically identified. Alternatively, do population-level data "average over" individual variation in ways that cause population-level patterns to converge too strongly to be able to statistically discriminate between models. Using simulation and machine learning classifiers, I examined whether it was possible to distinguish *different* mixtures of social learning strategies with only population-level summary data, in the form of relative frequencies of artifact classes.

The results were not encouraging, if population level data are also subject to sample size effects and any temporal aggregation. It proved possible in some cases to distinguish between mixtures of strategies and a population of the same size practicing unbiased copying, when the frequencies used were a complete census of the population, and when no aggregation was present. With partial samples of the population, equifinality between models (even under the ideal circumstances of a simulation experiment) increased. Similarly, when time averaging of observations is a factor, equifinality between model comparisons was strong, and with both factors, we are simply unable to distinguish between theoretical models with population-level data on trait frequencies. This does not bode well for identification of microevolutionary models of transmission in most archaeological situations.

8.1.1.3 A Systematic Method for Measuring Equifinality and Underdetermination

In addition to examining the limits of empirical sufficiency given population level data, Chapter 3 employed a statistical method for detecting equifinality among theoretical models which has considerable generality. The method simultaneously addresses theoretical overlap in outcomes, as well as the ways in which the spatiotemporal and sampling characteristics of our data may affect empirical sufficiency. I use the idea that distinguishable hypotheses (models) will generate predictive data distributions which can be separated by the *Bayes classifier* for the model set; sets of models which are equifinal given a set of variables and data collection treatments will not be separable.

The Bayes classifier is typically impossible to directly calculate, but it can be well approximated by using machine learning classifiers with sufficiently high model capacity (Hastie et al., 2009). In this study, I examined classifier performance in distinguishing pairs of transmission models, with samples from the same predictive data but with differing sample sizes and amounts of temporal aggregation. This combination of methods allows us to examine two questions simultaneously: (*a*) are a set of models or hypotheses distinguishable *even in theory*, given a set of observable variables, and (*b*) how do data collection treatments, or the empirical scale of our data in terms of resolution or duration, affect model identification? This method is more general than cultural transmission modeling, and in fact should be a standard question we ask about theoretical models and our data at the outset of a research project. As Perreault (2019) notes in his excellent recent book, far too many hypotheses in archaeology are empirically insufficient because they are underdetermined by any data we can feasibly collect.

8.1.1.4 Final Thoughts on the Microevolutionary Program

I started this dissertation work understanding that much of the cultural transmission modeling happening in archaeology was aimed at synchronic, reconstructionist goals; for example, attempting to determine what cognitive biases a past population might have displayed. It should be clear that there is no good answer to that question even in theory. Are modern communities "conformist"? Novelty seeking? The answer has to be "no": individuals are, and vary in that tendency, and this means the frequency and prevalence of different social learning behaviors varies both historically, and situationally. I began this work with the goal of trying to examine larger scale ways of modeling cultural transmission, but I did wonder whether models like Boyd and Richerson's could be "scaled up" to gather diachronic, longer-term and larger-scale predictions (much as Kandler and Shennan 2013 proposed).

This kind of rescaling approach has been incredibly successful in other disciplines, such as statistical physics and physical chemistry. In order to rescale (or "coarse grain", to use the term most often employed by physicists) our cultural transmission models, we would progressively observe them at larger scales, "averaging" over variation at lower scales to make predictions at larger scales. The research in Chapter 3 originally involved such an approach: I wanted to see if an explicit renormalization technique, of the type pioneered by Leo Kadanoff (1966; 2000), would show that mixtures of social learning modes "average out" when viewed at progressively coarser scales. The basic concept for this kind of renormalization is pictured schematically in Figure 8.1. To the extent that individuals "balance" each other's biases and social learning modes, the net effect when we "zoom out" may look unbiased.

There is certainly an inability to distinguish between mixtures of social learning modes, as doc-

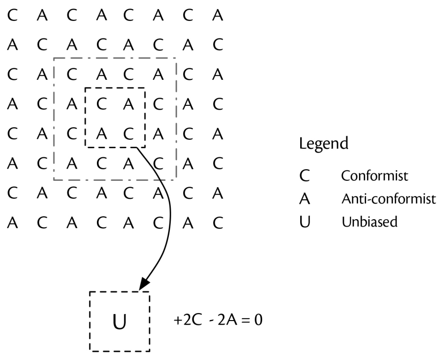


Figure 8.1: Example of "block renormalization" of a mixture of cultural transmission modes. Under the assumption that the conformist and anti-conformist biases are either equal and opposite, or drawn from random prior distributions, as one averages over more and more individuals, the population level effect begins to look more unbiased. This concept was the original impetus for the modeling approach taken in Chapter 3.

umented in Chapter 3, but I came to believe that an explicit renormalization approach to moving us from microevolutionary models to their mesoscopic consequences was not a fruitful one. The ability to renormalize a theory at a lower level and use it to predict consequences at larger scales turns out to depend upon two main properties, which are quite general. One may apply to cultural transmission modeling; the other definitely does not.

First, "coarse graining" a theoretical model requires *scale separation*, where some aspects of the detailed fine grained behavior of the system can be ignored at the higher level, because the effects become small enough to ignore (Hillerbrand, 2015). A classical example is using the "fine grained" theory of classical mechanics to predict the behavior of planets in our solar system (the coarse grained problem). We know that the "two body" problem is solvable, because the interaction between two objects still produces integrals with closed-form solutions. But even the "three body problem" (say, the Sun, Earth, and Moon) is not solvable. The degrees of freedom produce non-linear interactions and the system is no longer analytically integrable. Why then, can we still produce numerical answers with stunning accuracy? The contributions of some forces become exponentially small, given their size and distance, and thus can be ignored. Scale separation denotes the situation where some small scale effects become orders of magnitude smaller than the "main" effects in our models, and thus can be omitted. Whether scale separation applies if we "renormalize" transmission models incorporating cognitive biases is an open question; the answer may depend upon the relative strength of each effect. This is well worth investigating.

Second, usually in physical theory, the fine grained theory is "simpler" than the macro or mesoscale phenomena. Classical mechanics is time reversible and follows a few simple principles; it is the mesoscale phenomena like superconductivity, magnetism, and phases of matter that display complexity, phase transitions, and "critical points" where their behavior flips. The whole discipline of statistical physics aims, in fact, to understand universal principles for how combinations of simple rules give rise to complex and path-dependent behavior at large scales. As seductive as it is (at least to me!) to think that we could apply this kind of modeling paradigm in the social sciences, it seems increasingly clear to me that we are facing a different kind of situation.

The individual scale in social learning and cultural transmission theory is not simple—it is rich and variable, as much of anthropology, psychology, and other social sciences can attest. The situation we face is less like orbital mechanics and more like the study of turbulence in fluid dynamics, where small details lead to structure and pattern at all scales. And there is no good solution at the microscale for problems like turbulence. Instead, we study them empirically, modeling directly at the scales we need to, for the data we are attempting explain. The mathematical problem of studying turbulence directly at the "individual" or microscopic level remains out of reach—full generic solution of the Navier-Stokes equations remains an open Millennium Prize challenge, for example.

Thus, I have come to believe that we are in a situation where instead of employing individual-level models as explanations in themselves, our models need to directly model variation and history at mesoscopic and macroscopic scales. The strength of archaeology as an evolutionary discipline is time depth: we study a record of how human behavior differs over space and has changed over time. This requires us to construct models that are hypotheses about the evolutionary history of cultural variants in a region and over a period of time. Our evolutionary models should be hypotheses about the trajectories that different models of cultural transmission and social learning take, over time scales which match the time scales at which we can realistically observe the process.

8.1.2 Mapping Evolutionary History at the "Mesoscale" with Seriation Graphs

As we move away from microevolutionary analysis and towards historical, diachronic models at the "mesoscale," the modeling framework necessarily changes. Our hypotheses are no longer synchronic, equilibrium models but instead become historical narratives which describe why changes we see in artifact class frequencies occurred in the way they did. O'Hara (1988)'s distinction between "chronicle" and "history" provides the perfect framework for modeling this kind of evolutionary history. Chronicles are "the facts," while histories are explanations. Histories provide causal narratives that purport to explain the chronicle of empirical observations.

In Chapters 4, 5, and 6, I explored how to: (*a*) formalize a hypothesis about the regional history of community interaction and cultural transmission in the form of an " temporal network" model; (*b*) construct seriation graphs which summarize the spatial and temporal history of how cultural traits varied, and (*c*) develop methods for assessing the fit between a temporal network model and an empirical seriation graph, using summary statistics about seriation graph topology. In this structure, a set of interval temporal network models are the candidate "histories", and function as hypotheses. Seriation graphs function as one kind of evolutionary chronicle; in this case, it is a chronicle about the spatial and temporal change that we see in the frequency of different cultural traits. The topological properties of seriation graphs become the statistics we use in machine learning models to make judgments about equifinality, and to attempt to perform empirical fits to data.

In order for seriation graphs to function well as chronicles, they need to have rich enough state spaces that different causal histories will cause meaningful differences in the structure of the seriation solutions. This is an important reason that Lipo and I moved to a graph representation from our former work building multiple independent seriations out of empirical datasets (Lipo et al., 2015). But richness of structure means that seriations must incorporate many assemblages, not just a few. Small seriation solutions will nearly always underdetermine any set of candidate hypotheses, because small seriation solutions can take on only a limited number of "shapes" and thus there is little scope in the "state space" for the original data generating model to create distinguishable empirical effects. Thus, in Chapter 5, I worked with Carl Lipo on developing more efficient criteria for seriating large sets of assemblages. Our results with distance-minimization as the ordering criterion builds on earlier work by Kadane and Shepherdson, and in combination with our graph construction heuristics, yields the ability to easily analyze dozens of assemblages at a time and construct large seriation graphs that might display unique structure given different patterns of interaction.

Chapter 6 represents an initial attempt to put the pieces described above together, and determine whether it is possible, in theory, to discern hypotheses about regional transmission history, using seriation graphs as the empirical unit of observation. The results are positive and suggestive that further research would help outline the limits of the approach, and possibly improve discriminatory power between different classes of historical models. I return to the next steps for this research program in Sections 8.2.1 and 8.2.2.

8.1.3 Methods for Including Structured Information in Cultural Transmission Models

In addition to constructing better spatiotemporal models of cultural transmission at archaeological scales, we also need "thicker" descriptions (sensu Geertz 1973; Ryle 1949) of cultural variation and how social learning processes interact and coevolve with the *content* of cultural variation. Following ideas from Mesoudi and O'Brien (2008b), in Chapter 7 I considered how to build a model encapsulating the *dependency structure* between traits in the form of a graph. Simulated transmission dynamics were then subject to that dependency structure as well as social learning rules. Two types of social learning rule were then compared in a modified Axelrod model: individual innovation, as compared to direct instruction. The results show that, as one might expect, formal instruction or tutoring results in deeper, more complete cultural repertoires, in comparison to pure imitation and individual trial-and-error learning. This research originally appeared in a volume on the "learning hypothesis" for behavioral modernity among Upper Paleolithic hominids, and the results described here lend support to the idea that "behavioral modernity"—the explosion of complexity and variety seen in late Paleolithic artifact inventories—may be the result of the coevolution of new forms of social learning with cumulative technological development, rather than being a consequence of biological change or demographics.

At a methodological level, this work contributes new ways of modeling cultural transmission such that we can create credible hypotheses for how social learning rules and content can evolve together. This is necessary if we are going to move beyond formal, very abstract models for cultural transmission, and actually explain the evolutionary history of a technology. The effort to construct thicker models has been championed by Tostevin (2019) in archaeology, following the critically important work of William Wimsatt and colleagues (Wimsatt and Griesemer, 2007; Wimsatt, 2019). I discuss some possible future research directions in Section 8.2.3.

8.2 Next Steps and Future Research

The major goal of my dissertation research is to examine whether we can improve the empirical sufficiency of cultural transmission research in archaeology by: (*a*) understanding the sources of equifinality between models, and building methods to detect it; (*b*) modeling cultural transmission at spatial and temporal scales appropriate to the coarse grained nature of archaeological data, and (*c*) selecting empirical observable units (such as seriation graphs) which display enough variation between models, that our hypotheses are not underdetermined. The conclusions just described show, I believe, some success in achieving all three goals. Nevertheless, some of the approaches I employed, such as the use of temporal networks and seriation graphs in Chapter 6 remain in their infancy, with many questions still to answer and refinements possible. In this section, I describe next steps for several of the methods and approaches used in previous chapters.

8.2.1 Further Development of Temporal Network Models as Evolutionary Hypotheses

In Chapter 6, I examined the idea that we could represent a diachronic hypothesis about the history of cultural transmission in a region in the form of a so-called "temporal" or "time-varying" network. I proposed that interval temporal networks were a good tool for representing long term change in cultural transmission patterns between sedentary, nucleated communities. The temporal network model functions as a formalization of a standard "metapopulation" modeling framework, in which discernible subpopulations can be identified and sampled, and have sufficient persistence that one could (in theory) measure migration rates between the subpopulations. Not all of the empirical situations we wish to study as archaeologists fit into a metapopulation modeling strategy, however, and thus interval temporal networks will not be a good tool for representing evolutionary hypotheses in all situations. Interval networks would not be a good structural model for transmission in highly mobile populations or dispersed hunter gatherer communities, for example. Given how difficult it is to describe residential and community patterning in the deep Paleolithic record, we would need different observable models to build evolutionary chronicles for such situations. Such models may not be discrete, like graphs, but instead may be field-like, with gradients of class prevalence the primary objects of analysis. This remains an open question.

Within the class of problems for which interval temporal networks are appropriate, there are several avenues to follow and improve upon the results reported here. These include:

- Understanding how to characterize equivalence classes of temporal network structures, in the same way that we can order or classify static graphs;
- Describe equivalence classes of network structures that have similar behavior under transmission or diffusion processes.

The first element involves understanding good ways to characterize meaningful differences in interval graphs. In Chapter 6 I did not directly analyze the temporal network models themselves, instead looking for the *effect* their structure may have on transmission via seriations. This approach, derived from my earlier work on ways to characterize equifinality (see Chapter 3), is useful but highly indirect and computationally expensive. We can complement the approach I took here with a direct study of the interval temporal networks that form our transmission scenarios.

For static graphs, there are many results in graph theory that help characterize equivalence classes of graphs. In particular, functions of the Laplacian spectrum define classes of structures, as do the eigenvalues of the adjacency matrix. The "energy" of a graph (i.e., the sum of the eigenvalues of the adjacency matrix) appears to be strongly related to how the graph can be decomposed into subgraphs with differing structure (e.g., cycles, cliques, trees) (Estrada and Benzi, 2017; Gutman and Zhou, 2006). We also know, for example, We know that trees which have 4 or 5 distinct eigenvalues (at a given diameter) are completely characterized by their normalized Laplacian spectrum (Braga et al., 2015). That is, each distinct set of eigenvalues forms a distinct and distinguishable set of tree structures. The ability to understand structural equivalence classes is important for understanding when we should see differences in the *behavior* of a cultural transmission process across a set of network structures, since a large body of research has shown how graph topology affects diffusion-like processes on static graphs (see reviews in Castellano and Pastor-Satorras 2006; Durrett 2007; Grimmett 2018; Szabó and Fath 2007).

For time varying or temporal graphs, we are beginning to understand their structural properties (Nicosia et al., 2013, 2012), especially given strong interest by computer scientists in using temporal contact networks to study consensus algorithms, distributed systems design, and mesh network design. Similarly, given the importance of time varying contact networks in epidemiology, we are beginning to understand the behavior of diffusion processes across contact networks (Liu et al., 2013; Paré et al., 2017; Liu and Wang, 2014; Uribe et al., 2019; Santoro et al., 2011). Most of this work has focused on "contact" networks where edges represent short-duration events, so there is a need to examine the degree to which the behavior of diffusion processes on interval networks may differ from contact networks.

Finally, given a general understanding of how diffusion processes (including epidemics, cultural transmission, population genetics) operate on interval networks, we can return to the general approach taken in Chapter 6 and examine more systematically which transmission scenarios we can distinguish. To answer archaeological questions about "complex societies," for example, we would like to be able to formulate transmission scenarios where "hierarchy" exists. We also need to be able to model the transition from flatter "nearest neighbor" and "small world" patterns of connectivity to hierarchical patterns of interaction, in order to examine hypotheses about the origins of hierarchical forms of social organization, and its potential breakdown.

8.2.2 Statistical Properties of Seriation Graph Solutions and Transmission Scenarios

In order to perform statistical analysis on seriation solutions, including using classifier models to under equifinality, or clustering algorithms to understand similarities in structure, we need to extract summary statistics from seriation graphs. In Chapter 6 I focused on the "Laplacian spectrum": the set of eigenvalues and their multiplicities for the Laplacian matrix of a graph.[1] The Laplacian matrix for a general graph incorporates information about the pattern of edge connections through the adjacency matrix, and the density of connectivity (or "branchiness", for trees) through the degree matrix. I employed the eigenvalues of the Laplacian as a set of independent or predictor variables that I used to train a classifier model, measuring the ability to cleanly separate different graph structures and "predict" the true data generating model underlying the seriation graph.

Although this approach was successful, it is not clear how many classes of transmission scenarios can be distinguished. To understand how generic the approach might be, and what additional developments might improve it, I propose two research directions to further develop the method:

- We need to characterize how much variation in Laplacian spectra exists across seriation graphs with a given number of assemblages;
- We can explore richer representations for seriation relationships which will provide bigger state spaces for discriminating between hypotheses.

One thing that the approach taken in Chapter 6 did not clarify is the degree to which seriation graphs might not vary enough in their Laplacian spectra to distinguish between data generating processes. How large is the "state space" formed by possible eigenvalue spectra for all trees on N vertices? It may turn out that there is enough variation to differentiate transmission scenarios which are extremely different in their structure: for example, nearly regular graphs of the "nearest neighbor" type, and a lineage splitting model which contains separate components. What is unclear is whether seriation graphs are still "too similar" in structure such that they underdetermine more subtle com-

---

[1] Concurrent with my own work, Lewitus and Morlon (2016) employed Laplacian spectra to classify phylogenetic trees.

parisons between transmission scenarios, such as small world connections versus hierarchical social interaction.

A first step is to characterize how much the Laplacian spectra of trees with N vertices can vary. For example, although the number of uniquely labeled trees on, say, 20 assemblages is very large (Cayley's theorem yields 2.62×10^{23} distinct solutions), the number of distinguishable spectra may be much smaller given that trees have relatively limited variation in edge connectivity patterns. This is a purely quantitative question, and answerable using the tools of combinatorics and graph theory.

Second, it is possible to study other representations of seriation data as well which might contain more structure. Our IDSS seriation algorithm (Lipo et al., 2015) by default produces solutions which represent the "minmax" solution: the largest set of assemblages that can form a valid solution, linked by edges which represent the smallest differences between class frequencies among adjacent vertices (Figure 8.2). This "minmax" solution graph, whether constructed using unimodality as the ordering criterion or simple distance minimization (as described in Chapter 5), represents a kind of minimal spanning tree which accounts for both spatial and temporal effects on ordering assemblage frequencies.

That minimum spanning tree is constructed by *pruning* edges from an intermediate representation which is the union of all valid subsolutions (Lipo and I dubbed this the "sumgraph" or "aggregrate" graph representation). If the criterion being used is unimodality, the sumgraph will represent only those links where unimodal solutions could be found, and will be denser than a tree but sometimes sparser than the complete graph over N vertices. The sumgraph records more similarity information than a "minmax" solution tree, and can have loops, cliques, and richer overall structure. Because the graph structures are not strict trees, and can have richer connectivity patterns (which can vary in different areas of the solution), the Laplacian spectra of the intermediate sumgraph should exhibit more variation than the final minmax tree. This creates a large space of

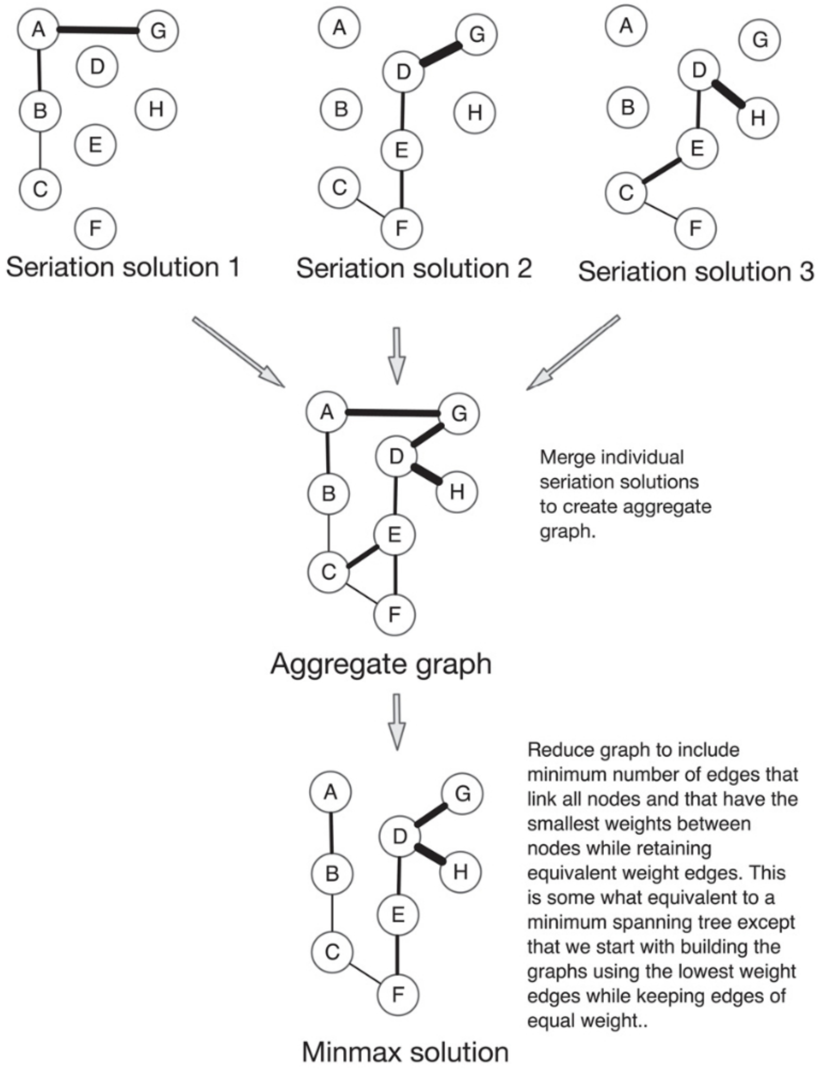


Figure 8.2: Seriation graph creation steps. In this example, we begin with the graph representation of three valid seriation solution fragments (1-3) for a set of 9 assemblages (A-I). In the figure, the thickness of the edges reflects the summed difference in frequencies between each pair of assemblages. Each solution represents a valid and unique seriation. To combine these three possible solutions into one overall solution, we first take the union of the partial graphs to create a single aggregate solution that is composed of all vertices and edges from the individual subsolutions. Using the aggregated graph or "sumgraph", we then reduce to a final seriation solution by including the fewest edges that can be made between all vertices and starting with the edges with the smallest weight (the sum of frequency differences). Edges are added sequentially ranked by total frequency difference until the vertices form a single connected three. Edges with equivalent total weight are retained. Reprinted from Figure 6, (Lipo et al., 2015) under the terms of the Creative Commons Attribution License `https://creativecommons.org/licenses/by/4.0/`.

variation within which we might distinguish between hypotheses in assessing equifinality, and potentially clearer ability to fit empirical data to different models. I have done some work repeating the experiments of Chapter 6 with sumgraphs rather than minmax seriation trees, and the results seem promising but there is a great deal that needs to be done to validate this line of reasoning.

8.2.3 Classification and Modeling Design Space

We always measure cultural variation, and map its spread through space and history through time, by examining the prevalence of archaeological classes of types (Dunnell, 1971). These classes are always analytical constructions, even if in practice archaeologists frequently employ standard classifications in a given region whose origins are a mixture of analysis and common sense. Since those classes are constructions, it is clear that "Baytown Plain" or "Elko corner-notched" are not really units of transmission, in the sense that the attributes denoted by those types were always learned or copied together. And yet, in much of the published literature on cultural transmission modeling in archaeology, we end up treating types as if they were units of transmission. This false equivalence causes us to ignore the actual levels at which cultural information is being transmitted—sometimes at the level of whole artifacts, sometimes at the level of individual attributes, and at levels in between given that individuals often learn *techniques* for construction and the attributes we observe are the byproducts of employing those techniques.

Instead, we should explicitly model the effects of classification and its "level" (fineness or coarseness) on our transmission models, in the same fashion that in Chapters 3 and 6 I explicitly modeled temporal aggregation in the transmission model itself. We can do this by modeling the artifact design space as a paradigmatic classification (Dunnell, 1971; O'Brien et al., 2015). Given a well constructed paradigmatic classification, we can vary the "fineness" with which we tabulate cultural variation by changing the "level" of the classification: adding or removing dimensions, or changing the gran-

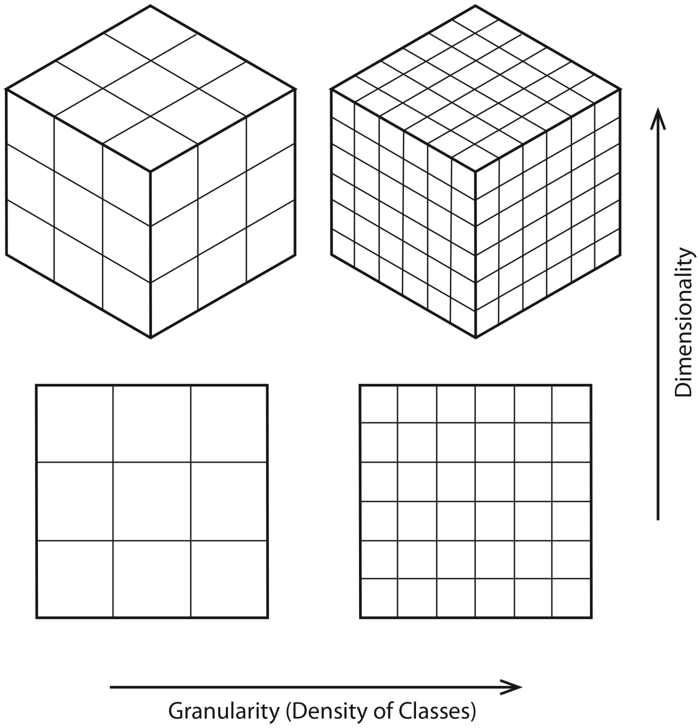


Figure 8.3: Factors which govern classification "level".

ularity with which we slice a dimension of variation into attributes or modes (see Figure 8.3). In our early work (Lipo et al., 1997), we changed the level of the Phillips et al. ceramic classification for the Lower Mississippi River Valley by lumping their types into larger aggregates, and performed seriations with the original and two levels of lumped classifications. When the resulting seriation solutions were mapped, the effect of varying the level of the classification is to create clusters of assemblages at different levels of spatial detail (Lipo et al., 1997, Fig. 17). Thus, changing the level of classification has the effect of revealing different levels of detail about the evolutionary history of variants, giving us a limited ability to "zoom in" or out to see pattern at different spatiotemporal scales.

Adding data collection treatments like assemblage sampling, time averaging, and variable classification level to a simple cultural transmission model adds a great deal of overhead and complexity.

In the research described in Chapter 6 I elected not to include classification level in the simulation models because the computational requirements of the model were already very challenging, and slowing things down even further would have led to even smaller numbers of replicates. But that is a software engineering problem, not a scientific one. I have examined how to add variable classificatory level to our simulation models, and extending the work in Chapter 6 to include it is a necessary next step.[2].

8.2.4 Next Steps on Modeling Structured Information

The vast bulk of research on cultural transmission modeling in archaeology has been concerned with: (*a*) what the frequency distribution of classes implies about selection or cognitive biases, and (*b*) what class frequencies can tell us about interaction in space and time. These problems, at least as originally framed, required only very "thin" descriptions of cultural variants. In other words, cultural variants in most of our models are simply markers, that possess counts and frequencies. This is to be expected, given the nature of the questions being asked. But answering questions about how technologies evolve through selection, social learning, and individual innovation require somewhat "thicker" descriptions in our models. As Gilbert Tostevin (2019) has argued, the *content* of cultural traits matter when constructing transmission models.

Modeling the content of cultural traits in terms of cultural transmission means recognizing that traits bear a variety of relationships to each other (Mesoudi and O'Brien, 2008b). Flake types in lithic assemblages relate to each other in terms of reduction sequences; the unpainted color of ceramic vessels is related to firing temperature and firing atmosphere, and so on. Attributes of technologies are related in terms of the steps needed to create a finished product, the prerequisite knowledge and skill

---

[2] A framework for tracking a simulation with a variable-level classification exists in the open-source repository `https://github.com/mmadsen/ctpy`, with a fuller Java-based implementation in `https://github.com/mmadsen/TransmissionFramework` that can be ported to Python for integration into the SimuPOP-based simulations described in this dissertation.

the artisan needs to perform a given technological transformation, and the locations, resources and other "scaffolding" needed to learn and execute it (Wimsatt and Griesemer, 2007; Wimsatt, 2019). Thicker descriptions are important in order to employ cultural transmission models to answer questions about the substantive history of technologies.

In Chapter 7 I attempted to model one type of dependency between traits: attributes that represent knowledge which is prerequisite to learning, and thus being able to execute, other attributes. I employed the tools of algebraic graph theory to examine the degree to which the pool of variation in observed assemblages fills the abstract "design space". I believe the framework has merit, but there are more relationships and dependencies that can and should be modeled. Premo and Tostevin (2016) describe how different classes might be observed, learned, and imitated in different ways in different locations, leading to their idea of the "taskscape". Traits or classes which are only observable in certain restricted contexts may exhibit more variability at a regional scale than those which are observable easily in many contexts, for example.

A good next step would be to examine whether the analysis from Chapter 7 could be done in the context of a model of lithic reduction sequences, and refine the symmetry analysis methods used there to examine how the "design space" for reduction sequence options fills in over time, and varies regionally. This kind of analysis would require, and lead to, deeper partnerships between specialists in different classes of prehistoric technologies, and specialists in evolutionary modeling.

But a word of caution is in order. As we develop more realistic models for dependencies between traits, we will need to be on our guard not to feel the "pull of the synchronic". There is merit in looking at how social learning interacts with the structure of a technology, but we should not fool ourselves into thinking that we are going to understand how flintknappers in the middle Paleolithic learned their craft, any more than we're going to learn how much Neolithic European potters were conformists or novelty-seekers. Our goal in thicker models for the transmission of technology should still be to understand change and build diachronic measures for change and innovation. A focus on classification, modeling of the design space and dependency relationships, and then building hypotheses about how design space is explored over time, should help keep our modeling efforts firmly at archaeological scales.

8.3 Final Thoughts

Cultural transmission, broadly understood, is the backbone of an evolutionary archaeology; through the principle of heritable continuity it provides the central organizing framework for the construction of evolutionary chronicles; and provides the mechanism through which other evolutionary processes act as part of our evolutionary explanations. Our exploration of cultural transmission theory and models in archaeology is, perhaps, out of its infancy, but principally because we have recently begun to explore its failures and potential limits, and respond with methodological and theoretical innovation to those issues.

Although I conclude in this work that evolutionary archaeologists should give up on the microevolutionary program given its conceptual and empirical weaknesses, the enterprise of building mathematical models of social learning and constructing statistical methods for fitting them to data remains absolutely critical. Those data will simply not be archaeological, in most cases. But we still need to understand how cultural transmission "works" as a system of heritability. Conclusions from theory building and studies at ecological scales can inform the theory building and studies we do at archaeological scales without being directly "renormalizable" between scales, or direct statistical fitting exercises. The relationship is instead much the same relationship that population genetics and evolutionary ecology have with paleobiology; that is, as sources of broad theoretical principles that are then turned into models and hypotheses of different kinds at different scales.

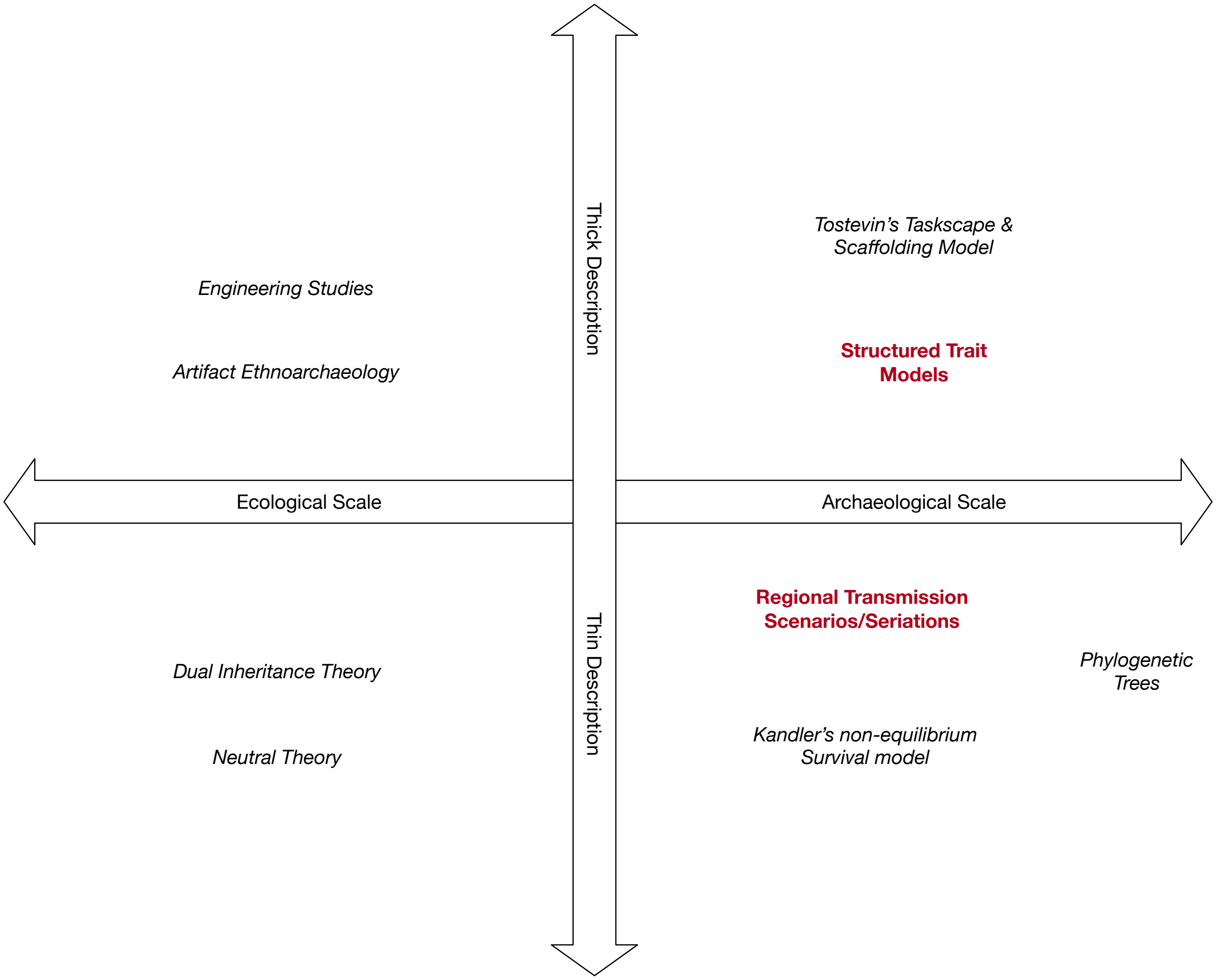


Figure 8.4: Conceptual relationship between the models and research discussed in this dissertation and cultural transmission studies within archaeology. The horizontal axis describes the time scale over which a body of theory of models are operative, from synchronic, "ecological time" models on the left, through diachronic, time averaged, "archaeological scale" models on the right. The vertical axis describes how "thin" or "thick" a model or theory is; does it treat cultural variants as abstract markers, or does it incorporate the "semantics" of how traits fit together: dependencies, engineering relationships, and prerequisites? The author's research in this dissertation is highlighted in red, to situate the work among other approaches.

In Figure 8.4, I attempt to depict the relationship between the various research programs and models discussed in this dissertation. The horizontal axis principally reflects time scale, from ecological scale studies of living populations, to archaeological scales which cover broad spans of space and time. The vertical axis reflects the degree to which cultural transmission models are "thin" and study just the statistical distribution of outcomes, or are "thicker" and model the actual content and relationships of cultural traits. The "mesoscale" research and models discussed in this dissertation are shown in red. The areas of theory and modeling given on the left side function as the synchronic science that research programs shown on the right can use to build diachronic chronicles and evolutionary explanations.

At the archaeological scale, we need both thin and thick efforts, for different purposes. The purpose of "thin" models at archaeological scales is to map the *flow* of cultural variation through space and time. Thin models give us the evolutionary chronicle of what happened when, where the innovations came from, to where did they spread, where they were adopted, where they were abandoned after they proved less than useful, and so on. Thick models take the engineering and functional information we derive from essentialist sciences like physics, chemisty, and various engineering disciplines, as well as rigorous ethnographic observations, to help us explain *how* technologies evolved, by allowing us to see our artifact classes (and their shifting frequencies) in the overall context of the technologies of which they are the phenotypic "hard parts".

On the right side of the diagram, I included several other bodies of research that are important for elaborating a truly archaeological scale cultural transmission theory. I have described the importance of Kandler's (2013) work on non-equilibrium, diachronic models elsewhere in this work, as well as Tostevin's (2012; 2019) work developing rich models for the social learning of lithic technology. Both pieces of work serve as exemplars of approaches we need to follow and further develop.

Phlyogenetic methods, and the macroevolutionary approach based upon them, have received less attention in this dissertation, but it is appropriate to understand their relationship to the "mesoscopic" work described here on seriation. I do not see seriation and cladistics as competitors or alternatives. They are, instead, ends of a continuum of methods for documenting the evolutionary chronicle. This continuum operates by varying *classificatory level* and by using categorical versus ratio scale variables. Although there has been some work on polymorphic phylogeny, employing frequencies of traits to understand finer-scale phylogenetic relationships (Wiens, 1999), the vast bulk of phylogenetic analysis employs the presence or absence of classes or traits (and their directionality of change, in cladistics) to determine relationships between taxa or samples. Seriation, both as classically employed and as rebuilt by Lipo and myself as an evolutionary method, can employ both presence/absence and frequencies, but the real value of the method comes in incorporating the rich frequency data that archaeological research generates.

Given these similarities and differences, it is typical that seriation is most appropriate when most assemblages or samples share the same group of classes (even if not all assemblages possess all classes), and where the differences arise in their frequency. Such situations are "zoomed in" often to relatively small intervals of time and small regions. This is the "mesoscale" as discussed in this dissertation. Phylogenetic methods are most appropriate when most assemblages or samples do not share the same group of classes, and we seek to understand evolutionary history simply through the pattern of their innovation and adoption. That is typically at larger spatial and longer time scales than questions we employ seriation to answer. The methods—and future derivatives of each, since there is much room for innovation here—are like selecting different lenses for a camera, rather than using different tools altogether. There will be much fruitful cross-pollination between the methods as well if we stop seeing them as alternatives or different methods and begin seeing them as a continuum of ways to create testable and secure evolutionary chronicles.

In the final analysis, the central task of an evolutionary archaeology is to provide the time depth and testable evolutionary history needed for the social sciences to study the richness of human history and lived experience in a scientific manner. Anthropology possesses the breadth of subject matter and time depth to lead this endeavor, partnering with the "ecological scale" disciplines like psychology and economics but providing the tools to handle diachronic change and historical contingency. Those tools are still under development, as my work here demonstrates. But the combination of century-old cultural historical methods with the quantitative methods of computational science, statistics, and various branches of mathematics is powerful, and we should look forward to rapid progress as the study of cultural transmission matures.

Bibliography

Abramowitz, M. and Stegun, L. 1965. *Handbook of Mathematical Functions*. Dover, New York.

Acerbi, A. and Bentley, R. A. 2014. Biases in cultural transmission shape the turnover of popular traits. *Evolution and Human Behavior*, 35(3):228–236.

Altman, D. G. 1991. *Practical statistics for medical research*. CRC Press.

Ammerman, A. J. and Cavalli-Sforza, L. L. 1971. Measuring the rate of spread of early farming in europe. *Man*, pages 674–688.

Antos, A., Devroye, L., and Gyorfi, L. 1999. Lower bounds for Bayes error estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(7):643–645.

Aoki, K. 2013. Determinants of cultural evolutionary rates. In Akazawa, T., Nishiaki, Y., and Aoki, K., editors, *Dynamics of Learning in Neanderthals and Modern Humans Volume 1*, Replacement of Neanderthals by Modern Humans Series, pages 199–210. Springer Japan.

Aoki, K. 2015. Modeling abrupt cultural regime shifts during the palaeolithic and stone age. *Theoretical population biology*, 100:6–12.

Aoki, K. and Feldman, M. 1987. Toward a theory for the evolution of cultural communication: coevolution of signal transmission and reception. *Proceeding of the National Academy of Sciences*, 84:7164–7168.

Aoki, K., Lehmann, L., and Feldman, M. W. 2011. Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission. *Theoretical population biology*, 79(4):192–202.

Aronica, G., Hankin, B., and Beven, K. 1998. Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Advances in Water Resources*, 22(4):349–365.

Arrow, K. 2009. Some developments in economic theory since 1940: An eyewitness account. *Annual Review of Economics*, 1(1):1–16.

Ascher, M. and Ascher, R. 1963. Chronological ordering by computer. *American Anthropologist*, 65(5):1045–1052.

Axelrod, R. 1997. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226.

Bailey, G. 1981. Concepts, time-scales and explanations in economic prehistory. In Sheridan, A. and Bailey, G., editors, *Economic archaeology: towards an integration of ecological and social approaches*, pages 97–118. British Archaeological Reports, International Series, no. 96.

Bailey, G. 1983. Concepts of time in quaternary prehistory. *Annual review of anthropology*, 12:165–192.

Bailey, G. 2007. Time perspectives, palimpsests and the archaeology of time. *Journal of Anthropological Archaeology*, 26(2):198–223.

Bailey, G. 2008. Time perspectivism: origins and consequences. *Time in archaeology: time perspectivism revisited*, pages 13–30.

Bailey, G. N. 1987. Breaking the time barrier. *Archaeological Review from Cambridge*, 6(1):5–20.

Bamforth, D. B. and Finlay, N. 2008. Introduction: Archaeological Approaches to Lithic Production Skill and Craft Learning. *Journal of Archaeological Method and Theory*, 15(1):1–27.

Banerjee, A. and Jost, J. 2008. On the spectrum of the normalized graph laplacian. *Linear algebra* *and its applications*, 428(11-12):3015–3022.

Bar-Yosef, O. 2002. The upper paleolithic revolution. *Annual Review of Anthropology*, pages 363–393.

Barrett, B. J. 2019. Equifinality in empirical studies of cultural transmission. *Behavioural processes*, 161:129–138.

Beals, R., Brainerd, G., Smith, W., Hack, J., and Jones, V. 1945. *Archaeological Studies in the Northeast Arizona*. University of California press.

Beaumont, M. A. 2010. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology*.

Beaumont, M. A., Zhang, W., and Balding, D. J. 2002. Approximate Bayesian computation in population genetics. *Genetics*.

Beineke, L. W., Wilson, R. J., Cameron, P. J., et al. 2004. *Topics in algebraic graph theory*, volume 102. Cambridge University Press.

Bentley, R. 2007. Fashion versus reason-then and now. *Antiquity*, 81(314):1071–1073.

Bentley, R., Hahn, M., and Shennan, S. 2004. Random drift and culture change. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1547):1443.

Bentley, R., Madsen, M., and Ormerod, P. 2009. Physical space and long-tail markets. *Physica A: Statistical Mechanics and its Applications*, 388(5):691–696.

Bentley, R. A., Lipo, C. P., Herzog, H. A., and Hahn, M. W. 2007. Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior*, 28:151–158.

Bentley, R. A. and Maschner, H. D. G. 2001. Stylistic change as a self-organized critical phenomenon: an archaeological study in complexity. *Journal of Archaeological Method and Theory*, 8(1):32.

Bentley, R. A. and Shennan, S. J. 2003. Cultural transmission and stochastic network growth. *American Antiquity*, 68(3):459–485.

Berl, R. E. and Hewlett, B. S. 2015. Cultural variation in the use of overimitation by the aka and ngandu of the congo basin. *PloS one*, 10(3).

Bertalanffy, L. v. 1969. General system theory: Foundations, development, applications.

Bettinger, R. and Eerkens, J. 1999. Point typologies, cultural transmission, and the spread of bow-and-arrow technology in the prehistoric great basin. *American Antiquity*, 64(2):231–242.

Bettinger, R. L. 2008. Cultural transmission and archaeology. In O'Brien, M., editor, *Cultural Transmsision and Archaeology: Issues and Case Studies*, pages 1–9. SAA Press.

Beven, K. 1996. Equifinality and uncertainty in geomorphological modelling. In *The Scientific Nature of Geomorphology: Proceedings of the 27th Binghamton Symposium in Geomorphology, Held 27–29 September 1996*, volume 27. John Wiley & Sons.

Beven, K. 2006. A manifesto for the equifinality thesis. *Journal of hydrology*, 320(1-2):18–36.

Billiard, S. and Alvergne, A. 2018. Stochasticity in cultural evolution: a revolution yet to happen. *History and philosophy of the life sciences*, 40(1):9.

Binford, L. 1981. Behavioral archaeology and the "pompeii premise". *Journal of Anthropological Research*, pages 195–208.

Binmore, K. 2005. *Natural justice*. Oxford University Press.

Bleed, P. 2001. Trees or chains, links or branches: conceptual alternatives for consideration of stone tool production and other sequential activities. *Journal of Archaeological Method and Theory*, 8(1):101–127.

Bleed, P. 2002. Obviously sequential, but continuous or staged? refits and cognition in three late paleolithic assemblages from japan. *Journal of Anthropological Archaeology*, 21(3):329–343.

Bleed, P. 2008. Skill Matters. *Journal of Archaeological Method and Theory*, 15(1):154–166.

Blitz, J. H. 2010. New perspectives in mississippian archaeology. *Journal of Archaeological Research*, 18(1):1–39.

Bluhm, E. 1951. Ceramic sequence in central basin and hopewell sites in central illinois. *American Antiquity*, 16:301–312.

Bonham, S. G., Haywood, A. M., Lunt, D. J., Collins, M., and Salzmann, U. 2009. El nino–southern oscillation, pliocene climate and equifinality. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1886):127–156.

Bordaz, V. v. H. and Bordaz, J. 1970. A computer pattern recognition method of classification and seriation applied to archaeological material. In Gardin, J.-C., editor, *Archéologie et Calculateurs*, pages 229–244. Centre National de la Recherche Scientifique.

Borgerhoff Mulder, M., Nunn, C. L., and Towner, M. C. 2006. Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 15(2):52–64.

Bouzouggar, A., Barton, N., Vanhaeren, M., d'Errico, F., Collcutt, S., Higham, T., Hodge, E., Parfitt, S., Rhodes, E., Schwenninger, J.-L., et al. 2007. 82,000-year-old shell beads from north africa and implications for the origins of modern human behavior. *Proceedings of the National Academy of Sciences*, 104(24):9964–9969.

Boyd, R. and Richerson, P. 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.

Braga, R. O., Del-Vecchio, R. R., Rodrigues, V. M., and Trevisan, V. 2015. Trees with 4 or 5 distinct normalized laplacian eigenvalues. *Linear Algebra and its Applications*, 471:615–635.

Braidwood, R. J. 1959. Archeology and the evolutionary theory. *Evolution and anthropology: a centennial appraisal*, pages 76–89.

Brainerd, G. W. 1951. The place of chronological ordering in archaeological analyis. *American Antiquity*, 16:301–312.

Breiman, L. 2001. Random forests. *Machine learning*, 45(1):5–32.

Broughton, J. M. and Grayson, D. K. 1993. Diet breadth, adaptive change, and the white mountains faunas. *Journal of Archaeological Science*, 20(3):331–336.

Brown, J. M. and Thomson, R. C. 2018. Evaluating Model Performance in Evolutionary Biology. *Annual Review of Ecology, Evolution, and Systematics*, 49(1):95–114.

Burnham, K. and Anderson, D. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Verlag.

Castellano, C., Fortunato, S., and Loreto, V. 2009. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591.

Castellano, C., Marsili, M., and Vespignani, A. 2000. Nonequilibrium phase transition in a model for social influence. *Physical Review Letters*, 85(16):3536.

Castellano, C. and Pastor-Satorras, R. 2006. Non-mean-field behavior of the contact process on scale-free networks. *Phys Rev Lett*.

Castro, L. and Toro, M. A. 2014. Cumulative cultural evolution: The role of teaching. *Journal of Theoretical Biology*, 347(0):74 – 83.

Cavalli-Sforza, L. and Feldman, M. 1973a. Cultural versus biological inheritance: phenotypic transmission from parents to children.(a theory of the effect of parental phenotypes on children's phenotypes). *American Journal of Human Genetics*, 25(6):618–637.

Cavalli-Sforza, L. and Feldman, M. 1973b. Models for cultural inheritance. I. Group mean and within group variation. *Theor. Popul. Biol.;(United States)*, 4(1).

Cavalli-Sforza, L. and Feldman, M. W. 1981. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press, Princeton.

Chang, L., Mak, M. C., Li, T., Wu, B. P., Chen, B. B., and Lu, H. J. 2011. Cultural adaptations to environmental variability: An evolutionary account of east–west differences. *Educational Psychology Review*, 23(1):99–129.

Chebotarev, P. 2013. Studying new classes of graph metrics. In *International Conference on Geometric Science of Information*, pages 207–214. Springer.

Chung, F. R. and Graham, F. C. 1997. *Spectral graph theory*. Number 92. American Mathematical Soc.

Cicchetti, D. and Rogosch, F. A. 1996. Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology*, 8(04):597–600.

Clauset, A., Shalizi, C., and Newman, M. 2007. Power-law distributions in empirical data. *Arxiv preprint arxiv:0706.1062*.

Cobb, C. R. 2003. Mississippian chiefdoms: How complex? *Annual Review of Anthropology*, 32(1):63–84.

Collard, M., Buchanan, B., Morin, J., and Costopoulos, A. 2011. What drives the evolution of hunter–gatherer subsistence technology? a reanalysis of the risk hypothesis with data from the pacific northwest. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1129–1138.

Collard, M., Buchanan, B., and OBrien, M. J. 2013a. Population size as an explanation for patterns in the paleolithic archaeological record. *Current Anthropology*, 54(S8):S388–S396.

Collard, M., Buchanan, B., O'Brien, M. J., and Scholnick, J. 2013b. Risk, mobility or population size? drivers of technological richness among contact-period western north american hunter–gatherers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1630):20120412.

Collard, M., Ruttle, A., Buchanan, B., and OBrien, M. J. 2013c. Population size and cultural evolution in nonindustrial food-producing societies. *PloS one*, 8(9):e72628.

Committee on Mathematical Foundations of Verification Validation and Uncertainty Quantification, National Research Council 2012. *Assessing the Reliability of Complex Models: Mathematical* *and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*. The National Academies Press.

Correa-Chávez, M. and Rogoff, B. 2009. Children's attention to interactions directed to others: Guatemalan mayan and european american patterns. *Developmental psychology*, 45(3):630.

Crane, H. et al. 2016. The ubiquitous ewens sampling formula. *Statistical science*, 31(1):1–19.

Craytor, W. B. and Johnson, L. 1968. *Refinements in computerized item seriation*. Museum of Natural History, University of Oregon.

Creanza, N., Fogarty, L., and Feldman, M. 2013. Exploring cultural niche construction from the paleolithic to modern hunter-gatherers. In Akazawa, T., Nishiaki, Y., and Aoki, K., editors, *Dynamics of Learning in Neanderthals and Modern Humans Volume 1*, Replacement of Neanderthals by Modern Humans Series, pages 211–228. Springer Japan.

Crema, E. R., Edinborough, K., Kerig, T., and Shennan, S. J. 2014. An Approximate Bayesian Computation approach for inferring patterns of cultural evolutionary change. *Journal of Archaeological Science*.

Crow, J. and Kimura, M. 1970. *An Introduction to Population Genetics Theory*. New York, Harper & Row.

Cruslock, E. M., Naylor, L. A., Foote, Y. L., and Swantesson, J. O. 2010. Geomorphologic equifinality: A comparison between shore platforms in höga kusten and fårö, sweden and the vale of glamorgan, south wales, uk. *Geomorphology*, 114(1-2):78–88.

Csibra, G. and Gergely, G. 2009. Natural pedagogy. *Trends in cognitive sciences*, 13(4):148–153.

Csibra, G. and Gergely, G. 2011. Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1149–1157.

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., and François, O. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, 25(7):410–418.

Culling, W. E. H. 1987. Equifinality: Modern Approaches to Dynamical Systems and Their Potential for Geographical Thought. *Transactions of the Institute of British Geographers*, 12(1):57.

De Sanctis, L. and Galla, T. 2009. Effects of noise and confidence thresholds in nominal and metric axelrod dynamics of social influence. *Physical Review E*, 79(4):046108.

Deetz, J. and Dethlefsen, E. S. 1965. The Doppler-effect and archaeology: a consideration of the spatial aspects of seriation. *Southwestern Journal of Anthropology*, 21:196–206.

Deetz, J. and Dethlefsen, E. S. 1971. Some social aspects of new england colonial mortuary art. *Memoirs of the Society for American Archaeology*, 25:30–38.

Demps, K., Zorondo-Rodríguez, F., García, C., and Reyes-García, V. 2012. Social learning across the life cycle: cultural knowledge acquisition for honey collection among the jenu kuruba, india. *Evolution and Human Behavior*, 33(5):460–470.

Dempsey, P. and Baumhoff, M. 1963. The statistical use of artifact distributions to establish chronological sequence. *American Antiquity*, pages 496–509.

Derex, M., Beugin, M.-P., Godelle, B., and Raymond, M. 2013. Experimental evidence for the influence of group size on cultural complexity. *Nature*, 503(7476):389–391.

d'Errico, F. and Henshilwood, C. S. 2007. Additional evidence for bone technology in the southern african middle stone age. *Journal of Human Evolution*, 52(2):142–163.

d'Errico, F. and Stringer, C. B. 2011. Evolution, revolution or saltation scenario for the emergence of modern cultures? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1060–1069.

Devijver, P. A. and Kittler, J. 1982. *Pattern recognition: A statistical approach*, volume 761. Prentice-Hall London.

Diestel, R. 2010. *Graph Theory, 4th Edition*. Springer-Verlag, Heidelberg.

DiNapoli, R. J., Lipo, C. P., Brosnan, T., Hunt, T. L., Hixon, S., Morrison, A. E., and Becker, M. 2019.

Rapa nui (easter island) monument (ahu) locations explained by freshwater sources. *PloS one*, 14(1).

Dobbin, K. K. 2009. A method for constructing a confidence bound for the actual error rate of a prediction rule in high dimensions. *Biostatistics (Oxford, England)*, 10(2):282–296.

Dunnell, R. C. 1970. Seriation method and its evaluation. *American Antiquity*, 35(3):305–319.

Dunnell, R. C. 1971. *Systematics in prehistory*. Free Press, New York.

Dunnell, R. C. 1978. Style and function: a fundamental dichotomy. *American Antiquity*, 43:192–202.

Dunnell, R. C. 1980. Evolutionary theory and archaeology. *Advances in archaeological method and theory*, 3:35–99.

Dunnell, R. C. 1981. *Seriation, groups, and measurements*, pages 67–90. Union International de Ciencias Prehistoricas y Protohistoricas, Mexico, DF.

Dunnell, R. C. 1982. The harvey lecture series. science, social science, and common sense: The agonizing dilemma of modern archaeology. *Journal of Anthropological Research*, pages 1–25.

Dunnell, R. C. 1986. Methodological issues in americanist artifact classification. In *Advances in archaeological method and theory*, pages 149–207. Elsevier.

Dunnell, R. C. 1989. Aspects of the application of evolutionary theory in archaeology. *Archaeological thought in America*, pages 35–49.

Durrett, R. 2007. *Random graph dynamics*. Cambridge Univ Pr.

Durrett, R. 2008. *Probability models for DNA Sequence Evolution*. New York, Springer, 2nd edition edition.

Ebel, B. A. and Loague, K. 2006. Physics-based hydrologic-response simulation: Seeing through the fog of equifinality. *Hydrological Processes: An International Journal*, 20(13):2887–2900.

Eerkens, J., Bettinger, R., and McElreath, R. 2006. Cultural transmission, phylogenetics, and the archaeological record. *Mapping our ancestors: Phylogenetic methods in anthropology and prehistory*, *ed. CP Lipo, MJ O’Brien, M. Collard & SJ Shennan*, pages 169–83.

Eerkens, J. and Lipo, C. 2005. Cultural transmission, copying errors, and the generation of variation in material culture and the archaeological record. *Journal of Anthropological Archaeology*, 24(4):316–334.

Eerkens, J. and Lipo, C. P. 2007. Cultural transmission theory and the archaeological record: Providing context to understanding variation and temporal changes in material culture. *Journal of Archaeological Research*, 15:239–274.

Efron, B. 1981. Nonparametric standard errors and confidence intervals. *canadian Journal of Statistics*, 9(2):139–158.

Efron, B. and Tibshirani, R. 1993. *An introduction to the bootstrap*, volume 57. Chapman & Hall/CRC.

Enquist, M., Eriksson, K., and Ghirlanda, S. 2007. Critical social learning: a solution to rogers's paradox of nonadaptive culture. *American Anthropologist*, 109(4):727–734.

Estrada, E. and Benzi, M. 2017. What is the meaning of the graph energy after all? *Discrete Applied Mathematics*, 230:71–77.

Evans, C. 1955. *A ceramic study of Virginia Archaeology*. BAE Bulletin 160, Washington.

Evans, T. S. and Giometto, A. 2011. Turnover rate of popularity charts in neutral models. *arXiv.org*, http://arxiv.org/abs/1105.4044.

Ewens, W. 1964. The maintenance of alleles by mutation. *Genetics*, 50(5):891–898.

Ewens, W. 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3(1):87–112.

Ewens, W. 1974. A note on the sampling theory for infinite alleles and infinite sites models. *Theoretical Population Biology*, 6(2):143–148.

Ewens, W. and Gillespie, J. 1974. Some simulation results for the neutral allele model, with interpretations. *Theoretical Population Biology*, 6(1):35–57.

Ewens, W. J. 2004. *Mathematical Population Genetics, Volume 1: Theoretical Introduction*. New York, Springer, 2nd edition.

Fehr, E. and Fischbacher, U. 2004. Third-party punishment and social norms. *Evolution and human behavior*, 25(2):63–87.

Feldman, M., Aoki, K., and Kumm, J. 1996. Individual versus social learning: evolutionary analysis in a fluctuating environment. *Anthropological Science*, 104:209–232.

Ferguson, J. R. 2008. The When, Where, and How of Novices in Craft Production. *Journal of Archaeological Method and Theory*, 15(1):51–67.

Flache, A. and Macy, M. W. 2006. What sustains cultural diversity and what undermines it? axelrod and beyond. *arXiv preprint physics/0604201*.

Fogarty, L., Strimling, P., and Laland, K. N. 2011. The evolution of teaching. *Evolution*, 65(10):2760–2770.

Ford, J. A. 1935. *Ceramic Decoration Sequence at an Old Indian Village Site Near Sicily Island, Louisiana*. Dept. Conservation, Louisiana Geological Survey, New Orleans.

Ford, J. A. 1936. *Analysis of Indian Village Site Collections from Louisiana and Mississippi*. Dept. Conservation, Louisiana Geological Survey, New Orelans.

Ford, J. A. 1938. A chronological method applicable to the southeast. *American Antiquity*, 3:260–264.

Ford, J. A. 1949. *Cultural dating of prehistoric sites in Viru Valley, Peru*, volume 43 of *Anthropological Papers*. American Museum of Natural History, New York.

Ford, J. A. 1962. *A Quantitative Method for Deriving Cultural Chronology*. Technical Manual, No. 1. Pan American Union.

Freund, Y. 1995. Boosting a weak learning algorithm by majority. *Information and computation*, 121(2):256–285.

Freund, Y., Schapire, R., and Abe, N. 1999. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612.

Friedman, J., Hastie, T., and Tibshirani, R. 2000. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2):337–407.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232.

Fukunaga, K. 1990. *Introduction to statistical pattern recognition*. Academic press.

Gardin, J.-C. 1970. A computer pattern recognition method of classification and seriation applied to archaeological material. Archchaeologie et Calculateurs:229–244.

Geertz, C. 1973. Thick description: Toward an interpretative theory of culture. In *The interpretation of cultures: selected essays*, pages 3–30. New York, Basic Books.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. 2013. *Bayesian data analysis*. CRC press.

Gelman, A., Meng, X.-L., and Stern, H. 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760.

Gelman, S. A. 2004. Psychological essentialism in children. *Trends in cognitive sciences*, 8(9):404–409.

Gifford-Gonzalez, D. 1991. Bones are not enough: analogues, knowledge, and interpretive strategies in zooarchaeology. *Journal of Anthropological Archaeology*, 10(3):215–254.

Gillespie, J. H. 1977. Sampling theory for alleles in a random environment. *Nature*, 266(5601):443–445.

Gintis, H. 2014. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences-Revised Edition*. Princeton University Press.

Gintis, H., Bowles, S., Boyd, R. T., Fehr, E., et al. 2005. *Moral sentiments and material interests: The foundations of cooperation in economic life*, volume 6. MIT press.

Gintis, H. et al. 2000. *Game theory evolving: A problem-centered introduction to modeling strategic behavior*. Princeton university press.

Godsil, C. D. and Royle, G. 2001. *Algebraic graph theory*, volume 8. Springer New York.

Gonzalez-Avella, J., Cosenza, M., and Klemm, K. 2007a. Information feedback and mass media effects in cultural dynamics. *Journal of Artificial Societies and Social Simulation*.

Gonzalez-Avella, J., Eguiluz, V., and San Miguel, M. 2007b. Homophily, Cultural Drift, and the Co-Evolution of Cultural Groups. *Journal of Conflict Resolution*.

González-Avella, J. C., Cosenza, M. G., and Tucci, K. 2005. Nonequilibrium transition induced by mass media in a model for social influence. *Physical Review E*, 72(6):065102.

González-Avella, J. C., Eguíluz, V. M., Cosenza, M. G., Klemm, K., Herrera, J., and San Miguel, M. 2006. Local versus global interactions in nonequilibrium transitions: A model of social dynamics. *Physical Review E*, 73(4):046119.

Grayson, D. and Delpech, F. 1998. Changing diet breadth in the early upper paleolithic of southwestern france. *Journal of Archaeological Science*, 25:1119–1129.

Grimmett, G. 2018. *Probability on graphs: random processes on graphs and lattices*, volume 8. Cambridge University Press.

Grote, M. N., Speed, T. P., et al. 2002. Approximate ewens formulae for symmetric overdominance selection. *The annals of applied probability*, 12(2):637–663.

Gutman, I. and Zhou, B. 2006. Laplacian energy of a graph. *Linear Algebra and its applications*, 414(1):29–37.

Hahn, M. W. and Bentley, R. A. 2003. Drift as a mechanism for cultural change: an example from baby names. *Proceedings of the Royal Society Biology Letters B*, 270:S120–S123.

Hamilton, M. J. and Buchanan, B. 2009. The accumulation of stochastic copying errors causes drift in culturally transmitted technologies: Quantifying clovis evolutionary dynamics. *Journal of Anthropological Archaeology*, 28:55–69.

Harary, F. 1969. *Graph Theory*. Addison-Wesley, Reading.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. 2009. *The elements of statistical learning*, volume 2. Springer.

Henrich, J. 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization*.

Henrich, J. 2006. Understanding cultural evolutionary models: A reply to read's critique. *American Antiquity*, 71:771–782.

Henrich, J. and Boyd, R. 2004. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the tasmanian case. *American Antiquity*, 69:197–214.

Herzog, H. A., Bentley, R. A., and Hahn, M. W. 2004. Random drift and large shifts in popularity of dog breeds. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 271:S353–S356.

Hillerbrand, R. 2015. Explanation via micro-reduction: On the role of scale separation for quantitative modelling. In *Why More Is Different*, pages 69–87. Springer.

Högberg, A. 2008. Playing with Flint: Tracing a Child's Imitation of Adult Work in a Lithic Assemblage. *Journal of Archaeological Method and Theory*, 15(1):112–131.

Holme, P. and Saramäki, J. 2012. Temporal networks. *Physics reports*, 519(3):97–125.

Huillet, T. 2007. Ewens sampling formulae with and without selection. *Journal of Computational and Applied Mathematics*, 206(2):755–773.

Hunt, T. D., Madsen, M. E., and Lipo, C. P. 1995. Examining cultural transmission using frequency seriation. In *Poster presented at the 60th SAA Annual Meeting, Minneapolis MN*.

Hunt, T. L. 1988. Graph theoretic network models for lapita exchange: a trial application. *Archae-*

ology of the Lapita cultural complex: A critical review, 5:135–156.

Irwin-Williams, C. 1977. A network model for the analysis of prehistoric trade. In *Exchange systems in prehistory*, pages 141–151. Elsevier.

Johnson, G. A. 1977. Aspects of regional analysis in archaeology. *Annual review of anthropology*, 6(1):479–508.

Jordan, P. and Shennan, S. 2003. Cultural transmission, language, and basketry traditions amongst the california indians. *Journal of Anthropological Archaeology*, 22(1):42–74.

Kadane, J. B. 1971. *Chronological Ordering of Archeological Deposits by the Minimum Path Length Method*. Center for Naval Analyses, Arlington, VA.

Kadanoff, L. 2000. *Statistical physics: statics, dynamics and renormalization*. World Scientific Pub Co Inc.

Kadanoff, L. P. 1966. Scaling laws for ising models near t c. *Physics Physique Fizika*, 2(6):263.

Kandler, A. and Crema, E. R. 2019. Analysing cultural frequency data: neutral theory and beyond. In *Handbook of Evolutionary Research in Archaeology*, pages 83–108. Springer.

Kandler, A. and Powell, A. 2018. Generative inference for cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1743):20170056.

Kandler, A. and Shennan, S. 2013. A non-equilibrium neutral model for analysing cultural change. *Journal of theoretical biology*, 330:18–25.

Kandler, A. and Shennan, S. 2015. A generative inference framework for analysing patterns of cultural change in sparse population data with evidence for fashion trends in lbk culture. *Journal of The Royal Society Interface*, 12(113).

Keeling, M. 2005. The implications of network structure for epidemic dynamics. *Theoretical Population Biology*, 67(1):1–8.

Keeling, M. and Rohani, P. 2007. *Modeling infectious diseases in humans and animals*. Princeton,

Princeton University Press.

Kempe, M. and Mesoudi, A. 2014. An experimental demonstration of the effect of group size on cultural accumulation. *Evolution and Human Behavior*.

Kendall, D. G. 1963. A statistical approach to flinders petrie's sequence dating. *Bulletin of the International Statistical Institute*, 40:657–680.

Kendall, D. G. 1969. Some problems and methods in statistical archaeology. *World Archaeology*, 1:68–76.

Kendall, D. G. 1970. A mathematical approach to seriation. *Philosophical Transactions of the Royal Society, Series A, Mathematical and Physical Sciences*, 269:125–135.

Kendall, D. G. a. 1971. Seriation from abundance matrices. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, pages 214–252.

Kendall, M. and Hill, A. 1953. The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society. Series A (General)*, 116(1):11–34.

Khatami, S., Peel, M. C., Peterson, T. J., and Western, A. W. 2017. Equifinality and process-based modelling. In *AGU Fall Meeting Abstracts*.

Khatami, S., Peel, M. C., Peterson, T. J., and Western, A. W. 2019. Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty. *Water Resources Research*, 55(11):8922–8941.

Khromov, P., Malliaris, C. D., and Morozov, A. V. 2018. Generalization of the ewens sampling formula to arbitrary fitness landscapes. *PLoS One*, 13(1):1–23.

Kidder, A. V. 1917. A design sequence from new mexico. *Proceedings of the National Academy of Sciences*, 3:369–370.

Kidwell, S. 1997. Time-averaging in the marine fossilrecord: Overview of strategies and uncertainties. *Geobios*, 30(7):977–995.

Kim, J.-H. 2009. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis*, 53(11):3735–3745.

Kimura, M. and Crow, J. 1964. The number of alleles that can be maintained in a finite population. *Genetics*, 49(4):725.

Kingman, J. 1977. The population structure associated with the ewens sampling formula. *Theoretical Population Biology*, 11(2):274–283.

Klein, R. G. 2009. *The human career: human biological and cultural origins*. University of Chicago Press.

Klemm, K., Eguíluz, V., Toral, R., and Miguel, M. 2003a. Global culture: A noise-induced transition in finite systems. *Physical Review E*.

Klemm, K., Eguíluz, V., Toral, R., and San Miguel, M. 2003b. Nonequilibrium transitions in complex networks: A model of social interaction. *Physical Review E*.

Klemm, K., Eguíluz, V., Toral, R., and Miguel, M. 2005. Globalization, polarization and cultural drift. *Journal of Economic Dynamics and Control*.

Kohler, T. A., VanBuskirk, S., and Ruscavage-Barz, S. 2004. Vessels and villages: evidence for conformist transmission in early village aggregations on the pajarito plateau, new mexico. *Journal of Anthropological Archaeology*, 23(1):100–118.

Kowalewski, M. 1996. Time-averaging, overcompleteness, and the geological record. *The Journal of Geology*, 104(3):317–326.

Krieger, A. D. 1944. The typological concept. *American Antiquity*, 3:271–288.

Kroeber, A. 1916. *Zuni Potsherds*. Anthropological Paper, No. 18, Part 1. American Museum of Natural History, New York.

Kroeber, A. 1937. Diffusion. *The Encyclopedia of Social Science, II*, pages 137–142.

Kroeber, A. L. 1923. *Cultural Patterns and Processes*. Harbinger, New York.

Kroeber, A. L. 1931. Historical reconstruction of culture growths and organic evolution. *American Anthropologist*, 33(2):149–156.

Kuhn, M. 2008. Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5):1–26.

Kuhn, M. and Johnson, K. 2013. *Applied predictive modeling*. Springer.

Kuhn, S. 2013. Cultural transmission, institutional continuity and the persistence of the mousterian. In Akazawa, T., Nishiaki, Y., and Aoki, K., editors, *Dynamics of Learning in Neanderthals and Modern Humans Volume 1*, Replacement of Neanderthals by Modern Humans Series, pages 105–113. Springer Japan.

Kuzara, R. S., Mead, G. R., and Dixon, K. A. 1966. Seriation of anthropological data: A computer program for matrix-ordering. *American Anthropologist*, 68(6):1442–1455.

Laland, K., Odling-Smee, J., and Feldman, M. 2000. Niche construction, biological evolution, and cultural change. *Behavioral and brain sciences*, 23(01):131–146.

Lamberg-Karlovsky, C. C. 1970. *Excavations at Tepe Yahya, Iran, 1967-1969: Progress Report*. Number 27. American School of Prehistoric Research, Harvard University.

Lanchier, N. 2012. The Axelrod model for the dissemination of culture revisited. *The Annals of Applied Probability*, 22(2):860–880.

Lanchier, N., Deijfen, M., Häggström, O., and Connor, S. 2010. Opinion dynamics with confidence threshold: an alternative to the Axelrod model. *Alea*.

Landau, D. and Binder, K. 2005. *A Guide to Monte Carlo Simulations In Statistical Physics*. Cambridge, Cambridge Univ Press.

Lehmann, L., Wakano, J. Y., and Aoki, K. 2013. On optimal learning schedules and the marginal value of cumulative cultural evolution. *Evolution*, 67(5):1435–1445.

Lewis, D. 1969. *Convention: A philosophical study*. Harvard University Pres, Cambridge.

Lewitus, E. and Morlon, H. 2016. Characterizing and comparing phylogenies from their laplacian spectrum. *Systematic biology*, 65(3):495–507.

Lewontin, R. C. 1974. Darwin and mendel—the materialist revolution. *The heritage of Copernicus: Theories "pleasing to the mind*, pages 166–183.

Lipo, C. P. 2001a. Community structures among late mississippian populations of the central mississippi river valley. In Hunt, T., Lipo, C. P., and Sterling, S., editors, *Posing Questions for a Scientific Archaeology*, pages 175–216. Bergin and Garvey, Westport.

Lipo, C. P. 2001b. *Science, Style and the Study of Community Structure: An Example from the Central Mississippi River Valley*. British Archaeological Reports, International Series, no. 918, Oxford.

Lipo, C. P. 2005. The resolution of cultural phylogenies using graphs. In Lipo, C., O'Brien, M. J., Collard, M., and Shennan, S. J., editors, *Mapping our ancestors: Phylogenetic approaches in anthropology and prehistory*. Transaction Publishers.

Lipo, C. P. 2006. *Mapping our ancestors: Phylogenetic approaches in anthropology and prehistory*. Aldine De Gruyter.

Lipo, C. P. and Eerkens, J. W. 2008. Culture history, cultural transmission, and explanation of seriation in the southeastern united states. In *Cultural Transmission and Archaeology: Issues and Case Studies*, pages 120–131. Society for American Archaeology Press, Washington, DC.

Lipo, C. P. and Madsen, M. E. 1997. The method seriation: Explaining the variability in the frequencies of types. In *62nd Annual Meeting for the Society for American Archaeology*.

Lipo, C. P. and Madsen, M. E. 2001. Neutrality, "style," and drift: building methods for studying cultural transmission in the archaeological record. In Hurt, T. D. and Rakita, G. F. M., editors, *Style and Function: Conceptual Issues in Evolutionary Archaeology*, pages 91–118. Bergin and Garvey, Westport, Connecticut.

Lipo, C. P., Madsen, M. E., and Dunnell, R. C. 2015. A theoretically-sufficient and computationally-practical technique for deterministic frequency seriation. *PLoS ONE*, 10(4):e0124942.

Lipo, C. P., Madsen, M. E., Dunnell, R. C., and Hunt, T. D. 1997. Population structure, cultural transmission, and frequency seriation. *Journal of Anthropological Archaeology*, 16(4):33.

Lipo, C. P., Madsen, M. E., and Hunt, T. D. 1995. Artifact style dynamics ii: Deriving seriation from a network model of transmission. In *Paper presented at the 60th Annual Meeting for the Society for American Archaeology, Minneapolis MN*.

Liu, Q. and Wang, X. 2014. Social learning in networks with time-varying topologies. *Asian Journal of Control*, 16(5):1342–1349.

Liu, S.-Y., Baronchelli, A., and Perra, N. 2013. Contagion dynamics in time-varying metapopulation networks. *Physical Review E*, 87(3):032805.

López, A., Correa-Chávez, M., Rogoff, B., and Gutiérrez, K. 2010. Attention to instruction directed to another by us mexican-heritage children of varying cultural backgrounds. *Developmental Psychology*, 46(3):593.

Lyman, R. 2003. The influence of time averaging and space averaging on the application of foraging theory in zooarchaeology. *Journal of Archaeological Science*, 30(5):595–610.

Lyman, R. 2004. The concept of equifinality in taphonomy. *Journal of Taphonomy*, 2(1):15–26.

Lyman, R. 2009. Graphing evolutionary pattern and process: a history of techniques in archaeology and …. *Journal of Human Evolution*.

Lyman, R. and Michael, J. 2003. *WC McKern and the midwestern taxonomic method*. University of Alabama Press, Tuscaloosa.

Lyman, R. and O'Brien, M. 2000a. Chronometers and units in early archaeology and paleontology. *American antiquity*, 65(4):691–707.

Lyman, R. and O'Brien, M. 2000b. Measuring and explaining change in artifact variation with clade-diversity diagrams. *Journal of Anthropological Archaeology*, 19(1):39–74.

Lyman, R. and O'Brien, M. 2001. The direct historical approach, analogical reasoning, and theory in americanist archaeology. *Journal of Archaeological Method and Theory*, 8(4):303–342.

Lyman, R. and O'Brien, M. 2006a. *Measuring time with artifacts*. University of Nebraska Press.

Lyman, R., O'Brien, M., and Dunnell, R. 1997. *The rise and fall of culture history*. Springer.

Lyman, R. L. 2008. Cultural transmission in north american anthropology and archaeology, ca. 1895–1965. *Cultural transmission and archaeology: Issues and case studies, ed. MJ O'Brien. Society for American Archaeology.*

Lyman, R. L. and O'Brien, M. 2006b. *Measuring Time with Artifacts*. University of Nebraska, Lincoln.

MacArthur, B. D., Sánchez-García, R. J., and Anderson, J. W. 2008. Symmetry in complex networks. *Discrete Applied Mathematics*, 156(18):3525–3531.

Madsen, M. and Lipo, C. P. 2015. An approach to fitting transmission models to seriations for regional-scale analysis. In *Paper presented at the 80th Annual Meeting of the Society for American Archaeology, San Francisco, CA*.

Madsen, M. E. 2012. Unbiased cultural transmission in time-averaged archaeological assemblages. *ArXiv e-prints*, 1204.2043.

Madsen, M. E. and Lipo, C. P. 2014. Combinatorial structure of the deterministic seriation method with multiple subset solutions. http://arxiv.org/abs/1412.6060.

Madsen, M. E., Lipo, C. P., and Bentley, R. A. 2008. Explaining seriation patterns through network-structured cultural transmission models. In *Poster presented at the 73rd Annual Meeting of the Society for American Archaeology*.

Mallios, S. 2014. Spatial seriation, vectors of change, and multicentered modeling of cultural transformations among san diego's historical gravestones: 50 years after deetz and dethlefsen's archaeological doppler effect. *Journal of Anthropological Research*, 70:69–106.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. 2012. Approximate bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180.

Marquardt, W. H. 1978. Advances in archaeological seriation. *Advances in Archaeological Seriation*, 1:257–314.

Marwick, B. 2005. What can archaeology do with boyd and richerson's cultural evolutionary program?

Matthews, J. 1963. Application of matrix analysis to archaeological problems. *Nature*, 198:930–934.

Mayer-Oakes, W. J. 1955. *Prehistory of the Upper Ohio Valley: A Introductory Study*. Carnegie Museum, Annals Vo. 34, Pittsburgh.

Mayr, E. 1959. Typological versus population thinking. *Evolution and Anthropology. Washington: The Anthropological Society of America*, pages 409–412.

McBrearty, S. 2007. Down with the revolution. *Rethinking the human revolution. Cambridge: MacDonald Institute for Archaeological Research Monographs*, pages 133–152.

McBrearty, S. and Brooks, A. S. 2000. The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of human evolution*, 39(5):453–563.

McElreath, R. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

McKay, B. D. and Piperno, A. 2014. Practical graph isomorphism, {II}. *Journal of Symbolic Computation*, 60(0):94 – 112.

McLachlan, G. J. 1975. Confidence Intervals for the Conditional Probability of Misallocation in Discriminant Analysis. *Biometrics*, 31(1):161.

Meggers, B. J. and Evans, C. 1957. *Archaeological investigation in the mouth of the Amazon*. Bureau of American Ethnology, Bulletin 167, Washington.

Mesoudi, A. 2014. Experimental studies of modern human social and individual learning in an archaeological context: People behave adaptively, but within limits. In *Dynamics of Learning in Neanderthals and Modern Humans Volume 2*, pages 65–76. Springer.

Mesoudi, A., Chang, L., Dall, S. R., and Thornton, A. 2016. The evolution of individual and cultural variation in social learning. *Trends in ecology & evolution*, 31(3):215–225.

Mesoudi, A., Chang, L., Murray, K., and Lu, H. J. 2015. Higher frequency of social learning in china than in the west shows cultural variation in the dynamics of cultural evolution. *Proceedings of the Royal Society B: Biological Sciences*, 282(1798):20142209.

Mesoudi, A. and Lycett, S. J. 2009. Random copying, frequency-dependent copying and culture change. *Evolution and Human Behavior*, 30:41–48.

Mesoudi, A. and O'Brien, M. J. 2008a. The cultural transmission of great basin projectile-point technology i: an experimental simulation. *American Antiquity*, 73(1):3–28.

Mesoudi, A. and O'Brien, M. J. 2008b. The learning and transmission of hierarchical cultural recipes. *Biological Theory*, 3(1):63–72.

Mesoudi, A., Whiten, A., and Dunbar, R. 2006a. A bias for social information in human cultural transmission. *British Journal of Psychology*, 97(3):405–423.

Mesoudi, A., Whiten, A., Laland, K., et al. 2006b. Towards a unified science of cultural evolution. *Behavioral and Brain Sciences*, 29(4):329–346.

Mesoudi, A., Whiten, A., Mesoudi, A., and Whiten, A. 2008. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 363:3489–3501.

Miller-Atkins, G. and Premo, L. 2018. Time-averaging and the spatial scale of regional cultural differentiation in archaeological assemblages. *STAR: Science & Technology of Archaeological Research*, 4(1):12–27.

Mills, B. J. 2017. Social network analysis in archaeology. *Annual review of anthropology*, 46:379–397.

Mills, B. J., Clark, J., Huntley, D., Haas, W. R., McMahan, B., Ryan, S. C., Trowbridge, M., Shackley, M. S., Breiger, R. L., and Roberts, J. M. J. 2010. The southwest social networks project: Prehistoric dynamics in the southwest u.s. In *75th Annual Meeting of the Society for American Archaeology*. Paper prepared for the symposium, ”New Approaches in Regional Network Analysis”.

Moore, M. W. 2010. “grammars of action” and stone flaking design space. In Nowell, A. and Davidson, I., editors, *Stone tools and the evolution of human cognition*, pages 13–43. University of Colorado Press, Boulder.

Moran, P. 1958. Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, pages 60–71. Cambridge Univ Press.

Moran, P. 1962. *The statistical processes of evolutionary theory.* Clarendon Press; Oxford University Press.

Murphy, K. P. 2012. *Machine learning: a probabilistic perspective.* MIT press.

Muthukrishna, M., Shulman, B. W., Vasilescu, V., and Henrich, J. 2014. Sociality influences cultural complexity. *Proceedings of the Royal Society B: Biological Sciences*, 281(1774):20132511.

Nakahashi, W. 2013. Cultural evolution and learning strategies in hominids. In Akazawa, T., Nishiaki, Y., and Aoki, K., editors, *Dynamics of Learning in Neanderthals and Modern Humans Volume 1*, Replacement of Neanderthals by Modern Humans Series, pages 245–254. Springer Japan.

Natekin, A. and Knoll, A. 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7:21.

Neff, H. 1992. Ceramics and evolution. *Archaeological Method and Theory*, 4:141–193.

Neiman, F. D. 1990. *An Evolutionary Approach to Archaeological Inference: Aspects of Architectural Variation in the 17th Century Chesapeake.* PhD thesis, Yale University.

Neiman, F. D. 1995. Stylistic variation in evolutionary perspective: Inferences from decorative diversity and .... *American Antiquity*, 60(1):7–36.

Nelson, N. C. 1916. Chronology of the tano ruins, new mexico. *American Anthropologist*, 18:159–180.

Nickel, M., Tresp, V., and Kriegel, H.-P. 2011. A three-way model for collective learning on multi-relational data. In Getoor, L. and Scheffer, T., editors, *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, ICML '11, pages 809–816, New York, NY, USA. ACM.

Nicosia, V., Tang, J., Mascolo, C., Musolesi, M., Russo, G., and Latora, V. 2013. Graph metrics for temporal networks. In *Temporal networks*, pages 15–40. Springer.

Nicosia, V., Tang, J., Musolesi, M., Russo, G., Mascolo, C., and Latora, V. 2012. Components in time-varying graphs. *Chaos: An interdisciplinary journal of nonlinear science*, 22(2):023101.

Nishiaki, Y., Aoki, K., and Akazawa, T. 2013. Introduction. In Akazawa, T., Nishiaki, Y., and Aoki, K., editors, *Dynamics of Learning in Neanderthals and Modern Humans Volume 1*, Replacement of Neanderthals by Modern Humans Series, pages 1–3. Springer Japan.

O'Brian, M. and Lyman, R. 2000. *Applying Evolutionary Archaeology*. New York: Kluwer Academic.

O'Brien, M., Darwent, J., and Lyman, R. 2001. Cladistics is useful for reconstructing archaeological phylogenies: Palaeoindian points from the southeastern united states. *Journal of Archaeological Science*, 28(10):1115–1136.

O'Brien, M. and Lyman, R. 1998. *James A. Ford and the growth of Americanist archaeology*. Univ of Missouri Pr.

O'Brien, M. and Lyman, R. 1999a. *Seriation, stratigraphy, and index fossils: The backbone of archaeological dating*. Plenum Pub Corp.

O'Brien, M. and Lyman, R. 2003. Resolving phylogeny: Evolutionary archaeology's fundamental issue. *Essential Tensions in Archaeological Method and Theory*, pages 115–125.

O'Brien, M., Lyman, R., and Darwent, J. 2000. Time, space, and marker types: James a. ford's 1936 chronology for the lower mississippi valley. *Southeastern Archaeology*, pages 46–62.

O'Brien, M., Lyman, R., Glover, D., and Darwent, J. 2003. *Cladistics and archaeology*. University of Utah Press Salt Lake City, UT.

O'Brien, M., Lyman, R., and Leonard, R. 1998. Basic incompatibilities between evolutionary and behavioral archaeology. *American antiquity*, 63(3):485–498.

O'Brien, M., Lyman, R., Mesoudi, A., and VanPool, T. 2010. Cultural traits as units of analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1559):3797–3806.

O'Brien, M. J., Boulanger, M. T., Buchanan, B., Bentley, R. A., Lyman, R. L., Lipo, C. P., Madsen, M. E., and Eren, M. I. 2015. Design space and cultural transmission: Case studies from paleoindian eastern north america. *Journal of Archaeological Method and Theory*, pages 1–49.

O'Brien, M. J. and Lyman, R. L. 1999b. *Seriation, Stratigraphy, and Index Fossils. The Backbone of Archaeological Dating*. Kluwer Academic/Plenum, New York.

O'Brien, M. J. and Lyman, R. L. 2000. *Applying evolutionary archaeology: A systematic approach*. Springer.

O'Brien, M. J. and Shennan, S. 2010. *Innovation in cultural systems: Contributions from evolutionary anthropology*. MIT Press.

Odling-Smee, F., Laland, K., and Feldman, M. 2003. *Niche construction: the neglected process in evolution*, volume 37. Princeton Univ Pr.

Odling-Smee, J. 2007. Niche inheritance: A possible basis for classifying multiple inheritance systems in evolution. *Biological Theory*, 2(3):276–289.

O'Hara, R. J. 1988. Homage to clio, or, toward an historical philosophy for evolutionary biology. *Systematic Biology*, 37(2):142–155.

Olszewski, T. 1999. Taking advantage of time-averaging. *Paleobiology*, 25:226–238.

Olszewski, T. 2004. Modeling the influence of taphonomic destruction, reworking, and burial on time-averaging in fossil accumulations. *Palaios*, 19(1):39–50.

Olszewski, T. 2011. Remembrance of things past: modelling the relationship between species' abundances in living communities and death assemblages. *Biology Letters*, 8(1):131–134.

Olszewski, T. and West, R. 1997. Influence of transportation and time-averaging in fossil assemblages from the pennsylvanian of oklahoma. *Lethaia*, 30(4):315–329.

Osgood, C. 1951. Culture: Its empirical and non-empirical character. *Southwestern Journal of Anthropology*, 7:202–214.

Otter, R. 1948. The number of trees. *The Annals of Mathematics*, 49(3):583–599.

Paré, P. E., Beck, C. L., and Nedić, A. 2017. Epidemic processes over time-varying networks. *IEEE Transactions on Control of Network Systems*, 5(3):1322–1334.

Peng, B. and Kimmel, M. 2005. simupop: a forward-time population genetics simulation environment. *Bioinformatics*, 21(18):3686–3687.

Peng, B., Kimmel, M., and Amos, C. 2012. *Forward-Time Population Genetics Simulations: Methods, Implementation, and Applications*. Hoboken, Wiley-Blackwell.

Perra, N., Baronchelli, A., Mocanu, D., Gonçalves, B., Pastor-Satorras, R., and Vespignani, A. 2012a. Random walks and search in time-varying networks. *Physical review letters*, 109(23):238701.

Perra, N., Gonçalves, B., Pastor-Satorras, R., and Vespignani, A. 2012b. Activity driven modeling of time varying networks. *Scientific reports*, 2:469.

Perreault, C. 2018. Time-averaging slows down rates of change in the archaeological record. *Journal of Archaeological Method and Theory*, 25(3):953–964.

Perreault, C. 2019. *The quality of the archaeological record*. University of Chicago Press.

Perreault, C. and Brantingham, P. 2011. Mobility-driven cultural transmission along the forager-collector continuum. *Journal of Anthropological Archaeology*, 30:62–68.

Petrie, F. W. M. 1899. Sequences in prehistoric remains. *Journal of the Anthropological Institute*, 29:295–301.

Pfeffer, M. T. 2001. The engineering and evolution of hawaiian fishhooks. In Hunt, T. L., Lipo, C. P., and Sterling, S., editors, *Posing Questions for a Scientific Archaeology*, pages 73–96. Bergin and

Garvey, Westport, Conn.

Phillips, P., Ford, J., and Griffin, J. 1951. *Archaeological Survey in the Lower Mississippi Alluvial Valley, 1940-1947*. Papers of the Peabody Museum of American Archaeology and Ethnology 25. Harvard University, Cambridge.

Pigliucci, M. and Müller, G. B. 2010. *Evolution: the extended synthesis*. MIT press.

Porčić, M. 2014. Exploring the effects of assemblage accumulation on diversity and innovation rate estimates in neutral, conformist, and anti-conformist models of cultural transmission. *Journal of Archaeological Method and Theory*, pages 1–22.

Premo, L. 2012. Local extinctions, connectedness, and cultural evolution in structured populations. *Advances in Complex Systems*, 15(01n02).

Premo, L. and Scholnick, J. 2011. The spatial scale of social learning affects cultural diversity. *American Antiquity*, 76(1):163–176.

Premo, L. S. 2010. Equifinality and explanation: the role of agent-based modeling in postpositivist archaeology. *Simulating Change: Archaeology into the Twenty-First Century. University of Utah Press, Salt Lake City*, pages 28–37.

Premo, L. S. 2014. Cultural Transmission and Diversity in Time-Averaged Assemblages. *Current Anthropology*, 55(1):105–114.

Premo, L. S. and Tostevin, G. B. 2016. Cultural transmission on the taskscape: Exploring the effects of taskscape visibility on cultural diversity. *PloS one*, 11(9).

Prentiss, A. M. and Laue, C. L. 2019. Cultural macroevolution. In *Handbook of Evolutionary Research in Archaeology*, pages 111–125. Springer.

Prentiss, A. M., Walsh, M. J., Foor, T. A., and Barnett, K. D. 2015. Cultural macroevolution among high latitude hunter–gatherers: a phylogenetic study of the arctic small tool tradition. *Journal of Archaeological Science*, 59:64 – 79.

Provine, W. 1989. *Sewall Wright and Evolutionary Biology*. Chicago, University of Chicago Press.

Provine, W. 2001. *The Origins of Theoretical Population Genetics*. Chicago, University of Chicago Press.

Rafferty, J., Neff, H., Fritz, G. J., Dunnell, R. C., Johnson, J. K., and Carr, P. J. 2008. *Time's River: Archaeological Syntheses from the Lower Mississippi Valley*. The University of Alabama Press.

Rhodes, M., Leslie, S.-J., and Tworek, C. M. 2012. Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, 109(34):13526–13531.

Rhodes, M. and Mandalaywala, T. M. 2017. The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(4):e1437.

Richerson, P. and Boyd, R. 2005. *Not by genes alone: How culture transformed human evolution*. University of Chicago Press.

Richerson, P. J. and Boyd, R. 2008. Response to our critics. *Biology & Philosophy*, 23(2):301–315.

Ridgeway, G. 1999. The state of boosting. *Computing Science and Statistics*, pages 172–181.

Rindos, D. 1984. *The origins of agriculture: an evolutionary perspective*. Academic Press.

Robert, C. P. 1994. *The Bayesian Choice A Decision Theoretic Motivation*. Springer Verlag.

Robinson, W. S. 1951. A method for chronologically ordering archaeological deposits. *American Antiquity*, 16(4):293–301.

Rogers, E. 2003. *The Diffusion of Innovations, 5th edition*. Free Press, New York.

Rorabaugh, A. N. 2014. Impacts of drift and population bottlenecks on the cultural transmission of a neutral continuous trait: an agent based model. *Journal of Archaeological Science*, 49:255–264.

Rothman, K., Greenland, S., and Lash, T. 2008. *Modern epidemiology*. Lippincott Williams & Wilkins.

Rotman, J. J. 1995. An introduction to the theory of groups, volume 148 of graduate texts in mathematics.

Rouse, I. 1967. Seriation in archaeology. In Riley, C. and Taylor, W., editors, *American Historical Anthropology*, pages 153–195. Southern Illinois University Press.

Rouse, I. B. 1939. *Prehistory in Haiti: A Study in Method*. Yale University Publications in Anthropology, No. 21, New Haven.

Rowe, J. H. 1959. Archaeological dating and cultural process. *Southwestern Journal of Anthropology*, pages 317–324.

Ryle, G. 1949. *The concept of mind*. Hutchinson University Library.

Santoro, N., Quattrociocchi, W., Flocchini, P., Casteigts, A., and Amblard, F. 2011. Time-varying graphs and social network analysis: Temporal indicators and metrics. *arXiv preprint arXiv:1102.0629*.

Savenije, H. H. G. 2001. Equifinality, a blessing in disguise? *Hydrological Processes*, 15(14):2835–2838.

Sawyer, S. and Hartl, D. 1985. A sampling theory for local selection. *Journal of Genetics*, 64(1):21–29.

Schank, R. C. and Abelson, R. P. 1977. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures (artificial intelligence series).

Schapire, R. E. and Freund, Y. 2012. *Boosting: Foundations and algorithms*. MIT Press.

Schiffer, M. 1987. *Formation Processes of the Archaeological Record*. University of New Mexico Press, Albuquerque.

Schiffer, M. B. 1983. Toward the identification of formation processes. *American Antiquity*, 48(4):675–706.

Scholnick, J. 2010. *Apprenticeship, Cultural Transmission and the Evolution of Cultural Traditions in Historic New England Gravestones*. PhD thesis, University of Arizona.

Scholnick, J. B. 2012. The spatial and temporal diffusion of stylistic innovations in material culture.

Advances in Complex Systems, 15(01n02).

Shennan, S. 2000. Population, culture history, and the dynamics of culture change1. *Current Anthropology*, 41(5):811–835.

Shennan, S. 2001. Demography and cultural innovation: a model and its implications for the emergence of modern human culture. *Cambridge Archaeological Journal*, 11(01):5–16.

Shennan, S. 2011. Descent with modification and the archaeological record. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1070–1079.

Shennan, S. and Wilkinson, J. 2001. Ceramic style change and neutral evolution: A case study from neolithic europe. *American Antiquity*, 66(4):577–593.

Shennan, S. J. and Bentley, R. A. 2008. Style, interaction and demography among the earliest farmers of central europe. SAA Press.

Shepardson, B. L. 2006. *Explaining Spatial and Temporal Patterns of Energy Investment In The Prehistoric Statuary of Rapa Nui (Easter Island)*. PhD thesis, University of Hawaií.

Shine, R., Brown, G. P., and Phillips, B. L. 2011. An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences*, 108(14):5708–5711.

Shott, M. 2008. Lower paleolithic industries, time, and the meaning of assemblage variation. *Time in Archaeology: Time Perspectivism Revisited. The University of Utah Press, Salt Lake City*, pages 46–60.

Sisson, S. A., Fan, Y., and Beaumont, M. 2018. *Handbook of approximate Bayesian computation*. Chapman and Hall/CRC.

Slatkin, M. 1994. An exact test for neutrality based on the ewens sampling distribution. *Genetical Research*, 64(01):71–74.

Slatkin, M. 1996. A correction to the exact test based on the ewens sampling distribution. *Genetical* *research*, 68(03):259–260.

Smilkov, D. and Kocarev, L. 2012. Influence of the network topology on epidemic spreading. *Physical Review E*, 85(1):016114.

Smith, C. S. 1950. *The archaeology of coastal New York*. American Museum of Natural History, Anthropological Papers 43(2), New York.

Smith, K. and Neiman, F. D. 2005. Frequency seriation, correspondence analysis, and woodland period ceramic assemblage variation in the deep south. *Southeastern Archaeology*, 26:49–72.

Spier, L. 1917. An outline for a chronology of zuni ruins. *Anthropological Papers of the American Museum of Natural History*, 18:209–331.

Steele, J., Glatz, C., and Kandler, A. 2010. Ceramic diversity, random copying, and tests for selectivity in ceramic production. *Journal of Archaeological Science*, 37(6):1348–1358.

Stein, J. 1987. Deposits for archaeologists. *Advances in archaeological method and theory*, 11:337–395.

Stein, J. K. 1993. Scale in archaeology, geosciences, and geoarchaeology. *Geological Society of America Special Papers*, 283:1–10.

Stein, J. K. 2001. A review of site formation processes and their relevance to geoarchaeology. In *Earth sciences and archaeology*, pages 37–51. Springer.

Stein, J. K., Deo, J. N., and Phillips, L. S. 2003. Big sites—short time: accumulation rates in archaeological sites. *Journal of Archaeological Science*, 30(3):297–316.

Sterelny, K. 2012. *The evolved apprentice*. MIT Press.

Stern, N. 1994. The implications of time-averaging for reconstructing the land-use patterns of early tool-using hominids. *Journal of Human Evolution*, 27(1-3):89–105.

Stern, N. 2008. Time averaging and the structure of late pleistocene archaeological deposits in south west tasmania. *Time in archaeology: Time perspectivism revisited*, pages 134–148.

Stout, D. 2002. Skill and cognition in stone tool production: An ethnographic case study from irian jaya 1. *Current Anthropology*, 43(5):693–722.

Stout, D. 2011. Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1050–1059.

Straus, L. G. 2005. A mosaic of change: the middle–upper paleolithic transition as viewed from new mexico and iberia. *Quaternary international*, 137(1):47–67.

Szabó, G. and Fath, G. 2007. Evolutionary games on graphs. *Physics reports*, 446(4-6):97–216.

Teltser, P. A. 1995. Culture history, evolutionary theory, and frequency seriation. In Teltser, P. A., editor, *Evolutionary Archaeology: Methodological Issues*, pages 51–68. University of Arizona Press, Tucson.

Tëmkin, I. and Eldredge, N. 2007. Phylogenetics and material cultural evolution. *Current Anthropology*, 48(1):146–154.

Terashima, H. 2013. The evolutionary development of learning and teaching strategies in human societies. In Akazawa, T., Nishiaki, Y., and Aoki, K., editors, *Dynamics of Learning in Neanderthals and Modern Humans Volume 1*, Replacement of Neanderthals by Modern Humans Series, pages 141–150. Springer Japan.

Terrell, J. 1977. Human biogeography in the solomon islands. *Fieldiana. Anthropology*, 68(1):1–47.

Tomasello, M. 2016. The ontogeny of cultural learning. *Current Opinion in Psychology*, 8:1–4.

Tomašových, A. and Kidwell, S. 2010a. The effects of temporal resolution on species turnover and on testing metacommunity models. *The American Naturalist*, 175(5):587–606.

Tomašových, A. and Kidwell, S. 2010b. Predicting the effects of increasing temporal scale on species composition, diversity, and rank-abundance distributions. *Paleobiology*, 36(4):672–695.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal* *of Royal Society Interface*, 6(31):187–202.

Tostevin, G. B. 2012. *Seeing lithics: a middle-range theory for testing for cultural transmission in the pleistocene*. Oxford: Oxbow Books.

Tostevin, G. B. 2019. Content matters. *Beyond the Meme: Development and Structure in Cultural Evolution*, 22.

Tumer, K. and Ghosh, J. 2003. Bayes error rate estimation using classifier ensembles. *International Journal of Smart Engineering System Design*, 5(2):95–109.

Tweed, R. G. and Lehman, D. R. 2002. Learning considered within a cultural context: Confucian and socratic approaches. *American Psychologist*, 57(2):89.

Uribe, C. A., Hare, J. Z., Kaplan, L., and Jadbabaie, A. 2019. Non-bayesian social learning with uncertain models over time-varying directed graphs. *arXiv preprint arXiv:1909.04255*.

Vapnik, V. 2013. *The nature of statistical learning theory*. Springer science & business media.

Villa, P. and Roebroeks, W. 2014. Neandertal Demise: An Archaeological Analysis of the Modern Human Superiority Complex. *PLoS ONE*, 9(4):e96424.

Vrugt, J. A., Ter Braak, C. J., Gupta, H. V., and Robinson, B. A. 2009. Equifinality of formal (dream) and informal (glue) bayesian approaches in hydrologic modeling? *Stochastic environmental research and risk assessment*, 23(7):1011–1026.

Wakano, J. and Aoki, K. 2007. Do social learning and conformist bias coevolve? henrich and boyd revisited. *Theoretical population biology*.

Wakano, J., Aoki, K., and Feldman, M. 2004. Evolution of social learning: a mathematical analysis. *Theoretical population biology*.

Wakeley, J. 2008. *Coalescent Theory*. Cambridge, Harvard University Press.

Walker, K. and Bambach, R. 1971. The significance of fossil assemblages from fine-grained sediments: time-averaged communities: Geological society of america annual meeting program with abstracts, v. 3.

Walsh, D. M. 2019. The paradox of population thinking: First order causes and higher order effects. *Evolutionary Causation: Biological and Philosophical Reflections*, page 227.

Walsh, M. J., Prentiss, A. M., and Riede, F. 2019. Introduction to cultural microevolutionary research in anthropology and archaeology. In *Handbook of Evolutionary Research in Archaeology*, pages 25–47. Springer.

Wandsnider, L. 2008. Time-averaged deposits and multitemporal processes in the wyoming basin, intermontane north america: A preliminary consideration of land tenure. *Time in Archaeology: Time Perspectivism Revisited*, page 61.

Watkins, J. 2010. Convergence time to the ewens sampling formula. *Journal of Mathematical Biology*, 60:189–206.

Watterson, G. 1974. The sampling theory of selectively neutral alleles. *Advances in Applied Probability*, pages 463–488.

Watterson, G. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical population biology*, 7(2):256–276.

Watterson, G. 1976. The stationary distribution of the infinitely-many neutral alleles diffusion model. *Journal of Applied Probability*, 13:639–651.

Watterson, G. 1978. The homozygosity test of neutrality. *Genetics*, 88(2):405–417.

Weibull, J. W. 1997. *Evolutionary game theory*. MIT press.

Whiten, A., Caldwell, C. A., and Mesoudi, A. 2016. Cultural diffusion in humans and other animals. *Current opinion in Psychology*, 8:15–21.

Whitlam, R. G. 1981. Problems in ceramic classification and chronology: An example from the mobile bay area, alabama. *Midcontinental Journal of Archaeology*, 6(2):179–206.

Wiens, J. J. 1999. Polymorphism in systematics and comparative biology. *Annual Review of Ecology* *and Systematics*, 30(1):327–362.

Wilcox, A. 1973. Indices of qualitative variation and political measurement. *The Western Political Quarterly*, 26(2):325–343.

Wilder, B. and Kandler, A. 2015. Inference of cultural transmission modes based on incomplete information. Forthcoming.

Wilhelmsen, K. H. 2001. Building the framework for an evolutionary explanation of projectile point variation: An example from the central mississippi river valley. In Hunt, T. L., Lipo, C. P., and Sterling, S., editors, *Posing Questions for a Scientific Archaeology*, pages 97–144. Bergin and Garvey, Westport, Conn.

Wimsatt, W. C. 2007. *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Harvard University Press.

Wimsatt, W. C. 2013. Articulating babel: An approach to cultural evolution. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4):563–571.

Wimsatt, W. C. 2014. Entrenchment and scaffolding: an architecture for a theory of cultural change. In Caporael, L. R., Griesemer, J. R., and Wimsatt, W. C., editors, *Developing Scaffolds in Evolution, Culture, and Cognition*, The Vienna Series in Theoretical Biology, pages 77–105. MIT Press.

Wimsatt, W. C. 2019. Articulating babel: An approach to cultural evolution. *Beyond the Meme: Development and Structure in Cultural Evolution*, 22.

Wimsatt, W. C. and Griesemer, J. R. 2007. Reproducing entrenchments to scaffold culture: The central role of development in cultural evolution. *Integrating evolution and development: From theory to practice*, pages 227–323.

Wissler, C. 1916. The application of statistical methods to the data on the trenton argillite culture. *American Anthropologist*, 18(2):190–197.

Wolpert, D. H. 2002. The supervised learning no-free-lunch theorems. In *Soft Computing and Industry*, pages 25–42. Springer.

Wolpert, D. H. and Macready, W. G. 1997. No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on*, 1(1):67–82.

Wright, S. 1931. Evolution in mendelian populations. *Genetics*, 16(2):97–159.

Yaveroglu, I. S. and Donthu, N. 2002. Cultural influences on the diffusion of new products. *Journal of International Consumer Marketing*, 14(4):49–63.

Vita

Mark E. Madsen

Department of Anthropology
Box 353100
University of Washington
Seattle, WA 98195 U.S.A.

Phone: 206-355-8477

EMAIL: mark@madsenlab.org

Born: October 22, 1966—Seattle, WA
Nationality: American

Current Positions

Chief Technology Officer and President, Ibis Networks, Honolulu, HI
Board Member and Past President, San Juan County Economic Development Council
President, Board of Trustees, San Juan Island Library District, Friday Harbor, WA
Board Member, Orcas Power and Light Cooperative, Eastsound, WA
Board Member and Past President, Board of Directors, Seattle Arts and Lectures, Seattle, WA

Areas of Specialization

Evolutionary theory and human culture; cultural transmission theory; social learning

Education

2012	PH.D. CANDIDACY in Anthropology, University of Washington, Seattle
1990	M.S. in Anthropology, University of Wisconsin-Madison
1988	B.A. CUM LAUDE, WITH COLLEGE AND DEPT. HONORS in Anthropology, University of Washington, Seattle

Work Experience

2016–	Ibis Networks, Honolulu, HI; *Chief Technology Officer and President*
2005–2006	Microsoft Corporation, Redmond, WA; *Senior Program Manager*
2002–2005	Network Clarity, Seattle, WA; *co-Founder & V.P., Product Development*

1997–2002 InterNAP Network Services, Seattle, WA; *Dir. of Software and Systems Engineering*
1995–1997 Realnetworks, Seattle, WA; *Software Engineer & Technical Webmaster*
1999–2005 Pinpoint Venture Group, Seattle, WA; *Managing Partner*
1995–2006 Emergent Media/AllRecipes.com, Seattle, WA; *co-Founder/President & Board Member*
1994–1995 U.W. Law School, Seattle, WA; *Systems Engineer*
1993–1994 Dept. of Anthropology, Seattle, WA; *Lab Manager*

Bibliography

JOURNAL ARTICLES

Lipo, Carl P., Mark E. Madsen, and Robert C. Dunnell (2015). "A Theoretically-Sufficient and Computationally-Practical Technique for Deterministic Frequency Seriation". In: *PLoS ONE* 10.4, e0124942. DOI: `10.1371/journal.pone.0124942`. URL: `http://dx.doi.org/10.1371%2Fjournal.pone.0124942`.

O'Brien, Michael J et al. (2015). "Design Space and Cultural Transmission: Case Studies from Paleoindian Eastern North America". In: *Journal of Archaeological Method and Theory*, pp. 1–49.

Madsen, Mark E. and Carl P. Lipo (2013). "Saving culture from selection: Comment on "An evolutionary framework for cultural change: Selectionism versus communal exchange" by L. Gabora". In: *Physics of Life Reviews* 10, pp. 149–150.

Madsen, Mark E. (2012b). "Unbiased Cultural Transmission in Time-Averaged Archaeological Assemblages". In: *ArXiv e-prints* 1204.2043. eprint: `http://arxiv.org/abs/1204.2043`. URL: `http://arxiv.org/abs/1204.2043`.

Bentley, R.A., M.E. Madsen, and P. Ormerod (2009). "Physical space and long-tail markets". In: *Physica A: Statistical Mechanics and its Applications* 388.5, pp. 691–696.

Kornbacher, Kimberly and Mark E. Madsen (1999). "Explaining the Evolution of Cultural Elaboration". In: *Journal of Anthropological Archaeology* 18.3, p. 2.

Madsen, Mark, Carl Lipo, and Michael D. Cannon (1999). "Fitness and Reproductive Trade-Offs in Uncertain Environments: Explaining the Evolution of Cultural Elaboration". In: *Journal of Anthropological Archaeology* 18.3, p. 31.

Hunt, Terry L., Mark E. Madsen, and Carl P. Lipo (1998). "Reply to Boone and Smith: Is it Evolution Yet?" In: *Current Anthropology* 39.1.

Lipo, Carl P. and Mark E. Madsen (1998). "Reply to Lyman and O'Brien: The Goals of Evolutionary Archaeology: History and Explanation". In: *Current Anthropology* 39.5, pp. 615–652.

Lipo, Carl P. et al. (1997). "Population structure, cultural transmission, and frequency seriation". In: *Journal of Anthropological Archaeology* 16.4, p. 33.

Lipo, Carl P. and Mark E. Madsen (1993b). "The Importance of Cultural Variability: An Evolutionary Perspective." In: *Bulletin of the National Association of Student Anthropologists* 8.3, pp. 3–5.

Madsen, Mark E. (1991). “Anthropology and the Conservation Movement”. In: *Bulletin of the National Assocation of Student Anthropologists* 7.2, pp. 12–14.

Madsen, Mark E. and Roger Lohmann (1991). “The Student’s Role in Projecting Anthropology to the Public as a Practical Endeavor”. In: *Anthropology Newsletter* 32.14–15.

Book Chapters and Edited Volumes

Madsen, Mark E. and Carl P. Lipo (2015a). “Behavioral Modernity and the Cultural Transmission of Structured Information: The Semantic Axelrod Model”. English. In: *Learning Strategies and Cultural Evolution during the Palaeolithic*. Ed. by Alex Mesoudi and Kenichi Aoki. Replacement of Neanderthals by Modern Humans Series. Springer Japan, pp. 67–83. isbn: 978-4-431-55362-5. doi: `10.1007/978-4-431-55363-2_6`. url: `http://dx.doi.org/10.1007/978-4-431-55363-2_6`.

Madsen, Mark E. (2001). “Evolutionary Bet-Hedging and the Hopewell Cultural Climax”. In: *Posing Questions for a Scientific Archaeology*. Ed. by T.L. Hunt, C.P. Lipo, and S. Sterling. Bergin and Garvey, pp. 279–304.

Lipo, Carl and Mark Madsen (2000a). “Neutrality, ”style,” and drift: Building methods for studying cultural transmission in the archaeological record”. In: *Style and Function: Conceptual Issues in Evolutionary Archaeology*. Ed. by Teresa D Hurt and Gordon F M Rakita. Westport, Connecticut: Bergin and Garvey, pp. 91–118.

Madsen, M.E. (1994). “Problems and Solutions in the Study of Dispersed Communities”. In: *Ohio Hopewell Community Patterning*. Ed. by W.S. Dancey and P.J. Pacheco. Kent State University Press, pp. 85–104.

– (1992b). “Lithic Manufacturing at British Camp: Evidence from Size Distributions and Microartifacts”. In: *Deciphering a Northwest Coast Shell Midden: Geoarchaeological Approaches to Site Formation*. Ed. by J.K. Stein. Academic Press.

Invited Lectures and Conference Papers

Lipo, C.P. and M.E. Madsen (2016). “Continuity-based approaches and the study of patterns of cultural inheritance”. In: *Paper presented at the 28th Annual Meeting of the Human Behavior and Evolution Society, Vancouver, WA*.

Madsen, M.E. (2016). “Computational Methods for Identifying Metapopulation Interaction Patterns From Seriation Solutions”. In: *Paper presented at the 28th Annual Meeting of the Human Behavior and Evolution Society, Vancouver, WA*.

– (March 2016). “The Science of Change: What Does It Take to Measure Cultural Transmission at Large Scales?” In: *Invited lecture in the EVoS Seminar Series, Binghamton University, Binghamton, NY*.

Madsen, M.E. and Carl P. Lipo (2016). “Measuring Cultural Relatedness Using Multiple Seriation Ordering Algorithms”. In: *Paper presented at the 81st Annual Meeting of the Society for American Archaeology, Orlando, FL*.

Madsen, M.E. and Carl P. Lipo (2015b). “An Approach to Fitting Transmission Models to Seriations for Regional-Scale Analysis”. In: *Paper presented at the 80th Annual Meeting of the Society for American Archaeology, San Francisco, CA*.

– (2015c). “Tools for Transparency and Replicability of Simulation in Archaeology”. In: *Paper presented at the 80th Annual Meeting of the Society for American Archaeology, San Francisco, CA*.

Madsen, Mark E. and Carl P. Lipo (2014). “Cultural Transmission of Structured Knowledge and Technological Complexity: The Axelrod Model Extended”. In: *Paper presented at the 79th Annual Meeting of the Society for American Archaeology, Austin TX*.

Madsen, Mark E. (2012a). “Improving the Fit Between Model and Data: Time-Averaging Effects on Unbiased Cultural Transmission”. In: *Paper presented at the 77th Annual Meeting of the Society for American Archaeology, Memphis TN*.

Madsen, Mark E, Carl P. Lipo, and R. Alexander Bentley (2008). “Explaining Seriation Patterns Through Network-Structured Cultural Transmission Models”. In: *Poster presented at the 73rd Annual Meeting of the Society for American Archaeology*.

Lipo, Carl P. and Mark E. Madsen (2001). “Organization”. In: *Paper presented at the 66th Annual meeting of the Society for American Archaeology, New Orleans LA*.

Madsen, Mark E., Carl P. Lipo, and Timothy D. Hunt (2001). “Cultural Elaboration, Life Histories, and Bet-Hedging Strategies in Unpredictable Environments”. In: *Paper presented at the 66th Annual Meeting of the Society for American Archaeology, New Orleans LA*.

Lipo, Carl P and Mark M Madsen (2000b). “Building Group Level Units in the Archaeological Record: An Example from the Mississippi River Valley”. In: *Paper presented at the 66th Annual Meeting of the Society for American Archaeology*.

Lipo, Carl P and Mark E. Madsen (1999). “Keywords in Evolutionary Archaeology: Neutrality”. In: *Invited paper presented at the 64th Annual Meeting of the Society for American Archaeology, Chicago*.

Madsen, Mark E. and Carl P. Lipo (1999). “Keywords in Evolutionary Archaeology: Fitness”. In: *Invited paper presented at the 64th Annual Meeting of the Society for American Archaeology, Chicago*.

Cannon, Michael D., Mark E. Madsen, and Carl P. Lipo (1998). “Wasteful Behavior in Evolutionary and Ecological Perspective”. In: *Paper presented at the 63rd Annual Meeting of the Society for American Archaeology, Seattle*.

Hunt, Terry L, Carl P. Lipo, and Mark E. Madsen (1998). “The Flaws of Phylogeny as History: A Pacific Islands Case Study”. In: *Paper presented at the 63rd Annual Meeting of the Society for American Archaeology, Seattle*.

Lipo, Carl P., Mark E. Madsen, and Michael D. Cannon (1998). “When Will “Wasteful” Behavior Occur?” In: *Paper presented at the 63rd Annual Meeting of the Society for American Archaeology, Seattle*.

Madsen, Mark E, Carl P. Lipo, and Michael D. Cannon (1998). “Building explanations for ’wasteful’ behavior in the archaeological record”. In: *Poster presented at the 63rd Annual Meeting of the Society for American Archaeology, Seattle*.

Lipo, Carl P. and Mark E. Madsen (1997). "The Method Seriation: Explaining the Variability in the Frequencies of Types". In: *62nd Annual Meeting for the Society for American Archaeology*.

– (1996). "Problems and solutions in the study of social organization". In: *61st Annual Meeting of the Society for American Archaeology*.

Hunt, Timothy D., Mark E. Madsen, and Carl P. Lipo (1995). "Examining Cultural Transmission Using Frequency Seriation". In: *Poster presented at the 60th SAA Annual Meeting, Minneapolis MN*.

Lipo, Carl P., Mark E. Madsen, and Timothy D. Hunt (1995). "Artifact Style Dynamics II: Deriving Seriation from a Network Model of Transmission". In: *Paper presented at the 60th Annual Meeting for the Society for American Archaeology, Minneapolis MN*.

Madsen, Mark E., Carl P. Lipo, and Timothy D. Hunt (1995). "Artifact Style Dynamics I: A Network Model of Cultural Transmission in Spatially Structured Populations". In: *Paper presented at the 60th Annual Meeting for the Society for American Archaeology, Minneapolis MN*.

Lipo, C.P. and M.E. Madsen (1993c). "Global Reductions of Diversity in Subsistence Systems: A Darwinian Argument for a Conservationist Anthropology". In: *Paper presented at the 92nd Annual Meeting of the American Anthropological Association*.

Madsen, Mark E. and Carl P. Lipo (1993). "Units of Evolution within a Darwinian Approach to Anthropology: Definitions and Implications for the Subfields". In: *Paper presented at the 1993 American Association of Anthropology meetings*.

Madsen, Mark E. (1992a). "Contemporaneity and the Analysis of Archaeological Settlement Patterning: An Example from the Ohio Hopewell". In: *Paper presented at the 57th Annual Meeting of the Society for American Archaeology, Pittsburgh PA*.

Lipo, Carl P. and Mark E. Madsen (1989). "Microartifacts and Macrointerpretations". In: *Paper presented at the 1989 South Asian Conference, Madison WI*.

Madsen, M.E. and R.C. Dunnell (1989). "The Role of Microartifacts in Deducing Land Use from Low Density Records in Plowed Surfaces". In: *Paper presented at the 54th Annual Meeting of the Society for American Archaeology*.

Madsen, Mark E. (1988a). "Lithic Manufacturing in a Northwest Coast Shell Midden: The Grain-size Evidence". In: *Paper presented at the Northwest Anthropological Conferences, Tacoma WA*.

– (1988b). "Microdebitage Analysis of a Northwest Coast Shell Midden". In: *Paper presented at the 53rd Annual Meeting of the Society for American Archaeology*.

Instructorship or TA Positions Held

Autumn 1994 Archaeology 205, Teaching Assistant to Dr. Donald Grayson, *University of Washington*

Grants, honors & awards

1989-1991	National Science Foundation Graduate Fellowship, University of Washington
1988	Regent's Fellowship, University of Wisconsin-Madison
1988-1990	Grants-in-Aid of Research, Sigma Xi Scientific Research Society
1988	Membership, Phi Beta Kappa National Honor Society