

Empirical Sufficiency and Neutral Theory: Building Seriation and Classification Into Archaeological Models of Cultural Transmission☆

Mark E. Madsen

Department of Anthropology, Box 353100, University of Washington, Seattle WA, 98195 USA

Abstract

Despite growing application of formal cultural transmission models in archaeology, the common approach is synchronic and reconstructionist, yielding few testable conclusions about the archaeological record. Culture-historical methods, especially seriation, are natural observational tools for fitting cultural transmission models to archaeological data, given that seriation is inherently diachronic and treats change as evolutionary and continuous. Continuing previous research, I construct and evaluate statistical models linking neutral theory, models of regional interaction, and seriation solutions. To do so, I employ a computational model of unbiased transmission within a regional metapopulation which explicitly embeds paradigmatic classification as the bridge between trait transmission and culture-historical method. I assess the utility of the methods developed using a case study from Carl Lipo's study of Late Prehistoric ceramics from the Phillips, Ford, and Griffin study area.

☆Revision: 72f09bc (2012-08-30)

Email address: mark@madsenlab.org (Mark E. Madsen)

URL: <http://madsenlab.org> (Mark E. Madsen)

Proposal Outline

1	Introduction	3
2	Case Study: Regional Interaction from Mississippian Ceramic Seriations	5
3	Previous Research	6
3.1	Issues in the Application of Cultural Transmission Models in Archaeology	6
3.2	Neutral Cultural Transmission Theory	9
3.2.1	Wright-Fisher Infinite-Alleles in a Single Population	9
3.2.2	Neutral Transmission in a Metapopulation	11
3.3	Lipo's Iterative Pairwise Seriation Method	13
4	Research Problems	14
5	Research Methods	16
5.1	Computational Modeling	16
5.1.1	Model Construction	16
5.1.2	Simulating Archaeological Classification	16
5.1.3	Implementing Metapopulation Cultural Transmission	18
5.2	Model Verification	19
5.2.1	Verification of Single-Population Unbiased Transmission	20
5.2.2	Verification of Metapopulation Transmission	21
5.3	Model Fitting and Analytic Methods	22
5.3.1	Strength of Evidence and Model Selection	23
6	Simulation Experiments: Research Questions 1-4	26
6.1	Question 1: Adequacy of Unbiased Cultural Transmission	26
6.2	Question 2: Behavior of Unbiased Transmission in Design Space	26
6.3	Question 3: What Metapopulation Dynamics Cause Multiple Seriation Solutions? . .	28
6.4	Question 4: What Drives Richness and Evenness Patterns in Seriation Solutions? . . .	29
7	Model Inference for PFG Ceramic Assemblages	30
7.1	Assemblages and Preliminary Analysis	30
7.2	Interaction Analysis	30
8	Estimated Timeline	31
9	Summary	32
	Appendix A Software Development Plan	41

1. Introduction

A key result of an evolutionary approach to archaeology is the realization that cultural transmission theory provides a grounding explanation for the success of culture-historical methods—and the seriation method in particular—in constructing chronologies and understanding regional-scale patterns in the archaeological record. Many of the early applications of formal transmission models in archaeology drew upon, and sought to enhance, traditional methods with the quantitative insights gained from theory.

Neiman (1995) introduced formal transmission models to archaeology in the form of the classic Wright-Fisher neutral theory, and used this model to study spatiotemporal patterns of interaction using ceramic types, given Dunnell's (1978) linkage of stylistic classes to homology and neutrality. Neiman further suggested that drift was sufficient to create the characteristic form of seriation solutions. Lipo, Hunt, Dunnell and this author (1997) employed seriation not to construct chronology but to map regional-scale patterns of interaction, by partitioning solutions into subsets of assemblages that seriate together successfully. Lipo (2001) extended this line of reasoning and deepened our understanding of the methods required to produce such analyses with real artifact assemblages.

In addition to grounding culture-historical methods in a mechanistic scientific theory, this line of research strongly suggests that *culture-historical methods are the natural observational tools for testing hypotheses that arise when considering cultural transmission processes as archaeological explanations*.

Although formal transmission models have become an important and growing area of archaeological research, the connection between culture-historical methods and transmission models has been largely ignored in recent years. Instead, the current approach treats cultural transmission models as if they directly describe archaeological data, in order to answer questions which are more behavioral and less evolutionary. Much of this body of work has focused upon identifying the “mode” of past cultural transmission, as if entire populations could easily be characterized as “conformist” or that modes of transmission were not continuously shifting and changing. The descriptions are often essentialist in units and lead to transformational explanations, despite nominally employing a theory of heritability and social learning which is non-essentialist and evolutionary in character.

At a methodological level, the difference between simple trait markers and the complex units we call “types” has not affected the models used in most studies, and the equilibrium or synchronic predictions made by cultural transmission models are treated as if they applied without modification to diachronic, time-averaged data derived from accretional deposits. The net result of this style of cultural transmission research in archaeology is that we have yet to produce many convincing fits of cultural transmission models to data, and thus few convincing explanations, as noted by Boyd and Richerson (2008). Ultimately, this style of analysis stems from taking—consciously or not—a synchronic, reconstructionist approach to applying cultural transmission theory to archaeology.

In contrast, the advantage of using seriation of analytic classes as the observational method for linking cultural transmission theory to archaeological data is that seriation treats change through time as continuous. Change is monitored through the frequencies of analytic artifact classes chosen to respond to variation over particular temporal and spatial scales. If we construct transmission models in such a way that their observable consequences are measured through seriations, then we avoid synchronic reconstructionism in applying cultural transmission theory to the archaeological record. A seriation *solution*, comprising a number of assemblages measured with a set of archaeological classes, is therefore the basic observational unit I intend to study in my dissertation research. How much variation in the richness and evenness of classes can we observe across seriation solutions, given different

45 models and parameterizations of cultural transmission models? If we form seriation solutions rigor-
46 ously, and break assemblages into groups which seriate cleanly together, but cannot be combined, what
47 aspects of regional-scale cultural transmission models govern the size distribution of such solutions?
48 The goal of my dissertation research is to answer such questions and develop seriation solutions as
49 a diachronic observational tool for the application of cultural transmission models to archaeological
50 data.

51 Development of seriation as a tool requires solving several methodological problems. Even with
52 innovations by Lipo and colleagues (Lipo et al. 1997; Lipo and Madsen 2000; Lipo 2001), three chal-
53 lenges remain. The first is “rewriting” neutral models to include observational units separate from the
54 “trait” information which flows within the population; observational units which represent the mul-
55 tidimensional nature of archaeological classes and types. The second challenge is matching the level
56 of modeling to the scale at which we have measured variability (not just central tendencies) in a given
57 empirical case. In many cases, assemblages are characterized by a single set of artifact class frequencies,
58 which means that descriptions of variability can only be obtained at the scale of multiple assemblages.
59 This requires models of cultural transmission within multiple populations, with the model structured
60 in a manner appropriate to the settlement and land-use patterns involved. The third need is devel-
61 opment of statistical models which link the structure and parameters of cultural transmission models
62 to variation in the quantitative properties of seriation solutions, when variation is measured using
63 archaeological classes and seriated using Lipo’s “iterative pairwise frequency seriation” method.

64 I propose to address these methodological and theoretical challenges in my dissertation research
65 through numerical simulation in two phases of modeling and analysis. The first phase of research
66 focuses upon understanding the dynamical and statistical behavior of cultural transmission models
67 augmented with archaeological classifications. The second phase of my research will consider the
68 methodological and practical issues involved in applying the result of Phase I to real archaeological
69 data. Carl Lipo’s (2001) dissertation research greatly expanded our initial foray into using seriation
70 to measure interaction between populations. In that work, he identified several clusters of ceramics
71 assemblages in the St. Francis and Memphis portions of the Phillips et al. (1951) study area that ap-
72 pear to represent strongly interacting populations, outside of which interaction was much less intense.
73 Given multiple seriation solutions (which display different average assemblage richness, and evenness
74 of classes represented), I ask *what model of information flow within a cultural metapopulation best ac-*
75 *counts for the observed pattern of seriation solution groups, and intra-seriation patterns of richness and*
76 *evenness?* Furthermore, I consider how we can perform statistical inference to select the model which
77 represents the best fit to specific seriations obtained in the course of an archaeological study.

78 In the next sections, I review issues with the current state of cultural transmission research in ar-
79 chaeology, and provide summaries of iterative frequency seriation and neutral cultural transmission
80 theory, in both its standard single population form, and across a metapopulation composed of in-
81 terconnected subpopulations. I then list detailed research questions, given the details of the models
82 and seriation method reviewed. Next, I describe the computational model I propose to construct to
83 address the research questions, and discuss its verification, testing, and model selection and statistical
84 methods. Finally, I describe specific research tasks that I propose for answering the questions outlined,
85 and provide an estimated plan and timeline for the work.

2. Case Study: Regional Interaction from Mississippian Ceramic Seriations

From 1940 through 1948, Philip Phillips, James A. Ford, and James B. Griffin made systematic collections of ceramics across the Lower Mississippi River Valley (Figure A.1).¹ The PFG study made collections from 383 different locations, producing over 346,000 ceramic samples Phillips et al. (1951). This monumental study firmly established the basic chronology of later prehistoric occupation in the Valley, demonstrated the utility of Ford's particular approach to ceramic seriation and chronology building, and gave rise to the basic culture-historical concept of "phase" as a space-time unit (Dunnell 1985; Lyman et al. 1997; O'Brien and Lyman 1998; Lipo et al. 1997; Lipo 2001).

For 46 assemblages collected by PFG in the St. Francis and Memphis areas, Lipo (2001) performed new seriations using a modification of Ford's deterministic frequency seriation technique. The original PFG seriations presented solutions for each analytical subdivision of the study area, and had departures from the unimodal expectation of the method Dunnell (1970). To yield solutions which meet the assumptions of the method, it is clear that not all assemblages will fit in the same seriation solution. Using error estimates for frequencies given binomial error terms to test whether differences in frequencies were significant, Lipo used an iterative approach to construct the largest seriation groupings possible from the original PFG solution. This type of seriation result I term a "*seriation solution group*," and this class of solutions are, I claim, the best archaeological observable for measuring the differences between differing models of cultural transmission within and between a set of archaeological assemblages, given diachronic and time-transgressive data. I describe Lipo's method, and additional significance tests which allow us great confidence in the final results, in greater detail in Section 3.3. The resulting seriation groups for the St. Francis and Memphis study areas are shown in Figure A.2.

Lipo's study area, given new fieldwork, comprised a portion of the original St. Francis and Memphis areas from the PFG study, but he also created seriation solution groups for the full set of assemblages in both PFG analytical areas (which met sample size requirements). These solutions, which are the result also of pairwise significance testing, are shown in Figure A.3. The spatial extent of each solution group is mapped in Figure A.4. Not only do these seriations yield a temporal order for assemblages in each spatial area, but immediately we can see that seriation groups as a whole differ in ways which can be related quantitatively to the statistical properties of an unbiased cultural transmission model. Seriation groups shown here differ in average class richness, and the evenness of frequencies. For example, Group 1 and Group 5, for example, have greater richness and a flatter diversity profile than Group 2b or Group 7.

Such differences arise in neutral theory principally through differences in the amount of new variation introduced into a population, either through endogenous innovation or the flow of information from outside a population (loosely, "migration").² Richness and evenness are measured in seriation with respect to a specific classification. These variables are thus always relative to a "design space" (O'Brien et al. 2010), and we need to understand the degree to which different instances of a neutral model (e.g., innovation rates and migration patterns) yield distinguishable values in that design space. Conversely, we need to understand the degree to which different instances of neutral models are equifinal when observed through a particular design space.

¹Hereafter, the study and its authors are referred to as "PFG" for brevity.

²I use the term "migration" and "migration rate" throughout given its prevalence in the theoretical literature, and am not implying residential relocation of people. Migration here denotes non-local information flow: individuals do move around a landscape, and have opportunities to spread information outside their own immediate social groups or local populations.

Therefore, the goal of my case study is to understand what *classes* of neutral models can best account for the histories seen in seriation solution groups both from Lipo's study area and for the remainder of the full PFG collection. Of the 383 assemblages PFG collected, only 220 assemblages have decorated sherds (the remainder being representing by Bell Plain, Neely's Ferry Plain, or both. Of the assemblages with decorated sherds, only 63 assemblages have more than 50 decorated sherds. I intend to use as many of these assemblages as possible (given consideration of sample size effects) to expand the seriation solutions created by Lipo (2001) and widen the area over which we understand the nature of interaction patterns in the Lower Mississippi River Valley.

3. Previous Research

3.1. *Issues in the Application of Cultural Transmission Models in Archaeology*

Since their introduction into archaeology by Neiman (1995), microevolutionary models of cultural transmission have become an increasingly important tool in studying the history and spread of artifact types. Transmission models and statistical methods borrowed from theoretical population genetics have been employed to study Paleoindian lithics (Hamilton and Buchanan 2009), basketry (Jordan and Shennan 2003), ceramics (e.g., Bentley and Shennan 2003; Steele et al. 2010; Lipo et al. 1997; Lipo and Madsen 2000; Lipo 2001; Shennan and Wilkinson 2001), and a variety of historical and contemporary data sets (e.g., Bentley 2005; Hahn and Bentley 2003; Bentley et al. 2009; Herzog et al. 2004; Bentley et al. 2007b; Scholnick 2010; Premo and Scholnick 2011).

Two main approaches are used by archaeologists in current applications of cultural transmission theory. The first is phylogenetic analysis, which employs cladistic methods for constructing hypotheses concerning homology from phenetic similarity (O'Brien et al. 2003). Cladograms are descriptions of patterns of descent, and do not contain information about the processes which gave rise to those patterns. Cladistic studies are often broadly "macroevolutionary," applied in archaeology to data over large regions or even continental-scale data sets. The second method is broadly "microevolutionary," confronting individual assemblages or sets of assemblages with detailed (often individual-scale) models of transmission processes, adopted and adapted from population genetics (and, to a lesser extent, epidemiology). The aim is to provide an "internal," finer-grained perspective upon the dynamics of cultural change by examining the events within a lineage that eventually lead to branching and phylogenesis.

As the experience of biology shows, a full picture of evolutionary history requires both, and progress in theory building and empirical explanation is often an interplay between micro and macroscopic methods and perspectives. In this review, however, I focus upon the issues that I see in microevolutionary modeling of cultural transmission—issues which I believe render many attempts to fit models to data both empirically and dynamically insufficient (Lewontin 1974).

As previously noted, microevolutionary applications of cultural transmission models are often synchronic and reconstructionist in structure. This manifests in several ways in published studies:

1. Modal characterization of cultural units
2. Employment of synchronic model predictions to compare to archaeological data
3. Treatment of diachronic, time-averaged deposits as if they represented synchronic units
4. Assumption that archaeological units represent actual units of past transmission

165 Modal characterization occurs whenever a study seeks to characterize past populations and ar-
166 chaeological deposits with only a central tendency of single description, ignoring variability over time
167 and space, and the heterogeneity that exists in all human populations. No anthropologist *intends* to
168 engage in modal characterization, and in our reflective moments everyone recognizes the existence
169 of variability and heterogeneity (if not always its causal role in evolutionary systems). But the study
170 of individual-level transmission rules in archaeological assemblages has been rife with modal charac-
171 terization, given attempts to link artifact class diversity profiles to “conformist” or other transmission
172 models (e.g., Kohler et al. 2004; Steele et al. 2010; Shennan and Wilkinson 2001).

173 Existing applications of formal models almost always employ synchronic or equilibrium model
174 predictions, as test statistics to compare with archaeological data. Clearly, understanding the equi-
175 librium or asymptotic properties of our mathematical models is the first step in comprehending the
176 possible behavior of an evolving system. But archaeologists face a record which is a *single realization*
177 of a complex, historical process, combining both predictable stochastic processes (such as unbiased
178 diffusion, or constant selection), and elements of chance (e.g., the unique order in which certain inno-
179 vations occur). The problem with focusing solely upon the stationary behavior of cultural transmission
180 models is that equilibrium states are *ahistorical*: many historical trajectories typically lead to an equi-
181 librium state in an “ergodic” system. If we wish to explain the actual phenomena of the archaeological
182 record, we need to focus on how evolutionary models generate *histories*. In a complex stochastic pro-
183 cess, this amounts to always asking, “how likely is a particular sequence of states,” rather than “which
184 equilibrium state does our data imply.”

185 Next, because of the tendency to slip back into a reconstructionist mode of explanation, we tend
186 to ignore the statistical and quantitative effects of aggregation and deposition upon the observations
187 we compare to theoretical models. Grayson and Delpech (1998) and Lyman (2003) make this point
188 forcefully in the realm of foraging theory and time averaged zooarchaeological assemblages. Archae-
189 ological deposits are almost always accretional palimpsests, representing cumulative artifact discard
190 over durations of varying length (Binford 1981; Schiffer 1987; Stein 1987; Bailey 2007, 1981). When
191 archaeologists measure the number of classes present in an assemblage, or calculate their relative fre-
192 quencies, such measures summarize the bulk properties of deposition and discard over a significant
193 span of time.³ Bentley et al. (2004) recognized this issue as presenting a problem for cultural trans-
194 mission applications, and claimed to deal with it across a series of studies by simply collapsing all of
195 their simulation runs into a single data set, with no analysis of the effects of variable duration. Ear-
196 lier, Neiman (1990) had argued that variable artifact discard rates had no effect on the frequencies
197 of traits within a population, as long as sufficient time averaging occurred to cancel out the effects of
198 discard rate variation. Nevertheless, in a recent study (Madsen 2012), I show that temporal aggrega-
199 tion of transmission events has strong quantitative effects which vary with assemblage duration and
200 innovation rate, causing statistical tests for neutrality to lose discriminatory power, and altering the
201 distribution of richness and evenness measures. This indicates that synchronic model predictions are
202 not good approximations for historical, aggregated data, and that we need to craft diachronic observ-
203 ables from our cultural transmission models to employ in studying archaeological phenomena.

204 Finally, there is a strong tendency in applications of cultural transmission models to treat the traits
205 we track though models as reflecting something “real” about units of transmission in past populations.
206 This can take extreme forms, when authors explicitly treat transmitted information as reflecting the

³As well as the action of various post-depositional and taphonomic processes, of course.

207 “mental template” that past people had for making artifact types, or it can lead to a concern that prob-
208 lems fitting models to archaeological data may result from not having the “correct” units with which
209 to measure transmission.⁴ Cultural transmission simply does not occur in predictable packages, and
210 it is futile to ask whether we can identify the “units” of transmission. Information is often transmit-
211 ted in a variety of packages and chunks, perhaps variable between individuals.⁵ And yet, using the
212 distinction between ideational (or analytical) units, and phenomenological units, we can measure the
213 *outcome* of transmission events without having to understand in detail the myriad of ways in which
214 individual bits of information flowed within a population (see esp. O’Brien et al. 2010). We do so by
215 employing paradigmatic classifications, designed to encompass the cultural variation we are studying
216 in a “design space,” with dimensions and modes applicable to variation at particular temporal and
217 spatial scales (Figure A.7).

218 In addition to these theoretical challenges, archaeological usage of cultural transmission models
219 has rarely considered the limitations of the statistical methods employed to test the fit between our
220 models and our data. The most common approach has been the treat neutrality as a “null model” or
221 hypothesis, against which processes of interest can be detected. Steele et al. (2010) correctly argue that
222 simply comparing an empirical to a theoretical frequency distribution is insufficient as a test of a com-
223 plex model, even when it is done using a theoretically informed statistical test. Moreover, they note,
224 many early studies have not even gone that far, and employ no tests, merely arguments of consistency
225 between a model and quantitative data (e.g., Bettinger and Eerkens 1999).⁶ Almost no studies employ
226 multiple test statistics or multiple models, and no study in the archaeological literature to date has
227 employed statistical model selection and a “strength of evidence” approach to finding the best trans-
228 mission model to account for the observed variability. Shennan describes our true challenge well in
229 his recent overview of cultural transmission studies in archaeology: “the real issue in most cases of
230 trying to understand the factors affecting variation in archaeological assemblages is less likely to be
231 the question of whether the variation is neutral or not, but what is the relative importance of various
232 selective and stochastic factors in accounting for it” (Shennan 2011, p. 1076). I agree and return to
233 these issues in Section 5.3.

234 In summary, these issues suggest that we need to reexamine the way we employ present cultural
235 transmission models in archaeology. We need to characterize variability, not just modal tendencies;
236 if we possess only a single set of measurements at a given scale (e.g., assemblages), then our analysis
237 will need to be regional, and characterize history and variability across a metapopulation, instead of
238 focusing upon modal characterization of single communities. We need to extract predictions from
239 models for diachronic, aggregated data, rather than attempting to “correct” synchronic predictions for
240 time averaging. And we need to separate the role of analytic classification in *measuring* the results

⁴An excellent example comes from a superb paper by Steele et al. (2010), where the authors conclude with concern about “refining our understanding of what features and categories can be treated as reliable units of prehistoric cultural transmission.”

⁵Nor is this really a differentiator with genetic and epigenetic transmission systems, reliant upon chemical coding systems. These systems of transmission certainly have a much more complex phenomenology than commonly recognized outside the biological sciences (e.g., Prohaska and Stadler 2008), with hierarchical information and flexible machinery for multiply interpreting nucleic acid sequences.

⁶I want to note that this does not imply that (Bettinger and Eerkens 1999) is a bad study. On the contrary, the authors did a superb job of analyzing the “package size” of cultural transmission for the transition to bow-and-arrow technology in two regions of the Great Basin. They simply did not perform any further statistical tests to validate their idea that differences in the package of information being transmitted were due to specific modes of transmission.

of past cultural transmission from the theoretical role played by information flowing through past populations. And we need better methods of evaluating the degree to which our data provide statistical evidence for or against alternative models.

3.2. *Neutral Cultural Transmission Theory*

In this section I review the main properties of unbiased cultural transmission models, both within a single population and within a metapopulation composed of interacting communities. There are several ways of modeling unbiased transmission in extended populations, each appropriate for explaining different empirical situations. For example, a continuous reaction-diffusion model would be the best choice for a mobile and dispersed population. For my dissertation research, the appropriate choice given my Late Prehistoric case study is a structured metapopulation model, given sedentary nucleated populations in the study area.

I focus in this project upon unbiased models of transmission, for several reasons. First, unbiased copying is the microevolutionary model which best reflects our intuition that homology is best traced using models which incorporate only copying and a drift component arising from stochastic sampling of a finite population over time (Dunnell 1978; Neiman 1995; Lipo and Madsen 2000). Second, since we expect real human populations to exhibit heterogeneity in the copying and imitation rules they employ, and for those rules to be chosen situationally or to shift over the human lifespan, it is likely that when averaged over an entire population with individuals at different stages of the life cycle, many of the transmission biases that are of interest when studying individual behavior will average themselves out. Also, given that we study accretional, diachronic deposits, time averaging (see Section 3.2.1) will further cause individual biases to combine and form an aggregate record which may be well described by an unbiased transmission model.

This second reason is, in fact, a conjecture, and I return to it in Research Question 1 below. If the above intuition is correct, then archaeologists may not need to concern themselves, when studying long-duration or regional scale data, with differences in “modes” of transmission or the fine details of dual-inheritance models.

3.2.1. *Wright-Fisher Infinite-Alleles in a Single Population*

Following Dunnell’s (1978) suggestion that the distribution of stylistic variation is explained by neutrality with respect to selection, Neiman (1995) introduced the Wright-Fisher infinite-alleles (WF-IA) model to archaeology as a formal description of unbiased cultural transmission within prehistoric populations. Here I review the main characteristics of WF-IA as employed by archaeologists.

The well-mixed neutral Wright-Fisher infinite-alleles model (Kimura and Crow 1964) considers a single dimension (“locus”) at which an unlimited number of variants (“alleles”) can occur, in a population of N individuals.⁷ 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).

⁷Conventionally, the model treats a diploid population, in which N individuals are composed of $2N$ 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 $4N\mu$.

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.⁸ 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.⁹ An important consequence of 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 .

The distribution of variants in the full population is complex, and exact expressions are usually unknown for WF-IA, with most population properties obtained by diffusion approximation (Ewens 2004). Archaeologists are fortunate, however, that the *sampling* properties of the model are well understood, since we always deal with samples of past human activity, rather than population-level census information. The basic result, due to Ewens (1972), is the joint distribution of variant counts in a sample of size n , which is given by the Ewens sampling distribution, given the population innovation rate (θ):

$$\mathbb{P}_{\theta,n}(a_1, \dots, a_n) = \frac{n!}{\theta^{(n)}} \prod_{j=1}^n \frac{(\theta/j)^{a_j}}{a_j!} \quad (1)$$

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 :

$$\mathbb{P}(n_1, n_2, \dots, n_k | k_{\text{obs}}) = \frac{n!}{|S_n^k| k! n_1 n_2 \cdots n_k} \quad (2)$$

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 the core probability distributions which compose the WF-IA model, many observable quantities have been calculated, including the form of the trait frequency spectrum which describes the expected “evenness” of traits at a given innovation rate, the expected richness of traits in a sample of given size, and various expected times to exit or fixation (in a non-infinite-alleles model). The core probability distributions also yield statistical tests by determining the likelihood that a sample of given size is drawn from Equation 2. 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).

These expectations apply at equilibrium to synchronic samples, and given the ergodic hypothesis,

⁸An 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.

⁹It is important to note that θ is not a measure of the “diversity” of traits in the population, as it has been employed in several archaeological studies, but is instead a *rate* parameter of the model.

to diachronic samples that constitute true time-averages (which are possible in contemporary field or laboratory settings). Accretional assemblages are not true “time-averages” in the sense usually employed in mathematics, statistics, and physics, but instead are *cumulative* samples over a duration. In Madsen (2012), I describe the results of numerical simulations designed to determine whether model expectations, or the power of statistical tests of neutrality, are modified when aggregated in the same manner as cumulative archaeological samples. The results show that once a sample duration exceeds the mean trait lifetime (not necessarily the mean lifetime of an archaeological type, but a modeled trait), measured richness begins to increase, diversity curves are flattened, and neutrality tests are subject to excessive Type I error.

3.2.2. Neutral Transmission in a Metapopulation

Cultural transmission in human groups is almost never constrained to occur within a single, closed group. Instead, even if people might be influenced more heavily by immediate family, friends, and neighbors, there is also considerable information exchange between groups and individuals who are unrelated or geographically distant. Exogamy, trade, and the migration of entire groups are some of the proximate mechanisms for information flows that are outside immediate social or residential groups. Thus, even if we can approximate the dynamics of neutral transmission within a group by the model described above, when we seek to describe patterns among groups, or across larger regions, we need models which explicitly contain population structure.

This can be done in several ways. If an empirical case involves highly mobile individuals, or groups that practice dispersed settlement systems, then continuous spatial models may be required (e.g. Kandler and Steele 2009; Kandler and Laland 2009). In situations where populations are sedentary and nucleated, in contrast, a metapopulation approach is warranted, and is the approach taken here given the Late Prehistoric case study described in Section 2. In the metapopulation approach, populations are spatially structured into local subpopulations (or demes), with migration among the demes which affects the local and often global dynamics of the processes being modeled. Often, not invariably, demes can come into being through colonization, and become extinct within a metapopulation (see papers in Hanski and Gilpin 1997, for an introduction to the diversity of metapopulation model applications).

Figure A.5 schematically depicts two metapopulation models involving four demes. Model A represents the simplest metapopulation model, originally introduced by Sewall Wright as the “island” model (Wright 1943), and much analyzed in theoretical population genetics. In the simplest model, neutral drift proceeds within each of the four demes (indicated by the looped arrows), and individuals are able to migrate between any pair of demes, at a constant rate for the whole metapopulation. This model is “well-mixed” in a regional sense, although it is structured into separate interacting populations at the individual level. This structure yields some interesting results. In general, the effective population size (N_e) is lower in a structured metapopulation than in a single population of comparable census size (Barton and Whitlock 1997).

In Model A, the variance in traits across local demes is given by Wright’s formula:

$$F_{st} \approx \frac{1}{2Nm + 1} \quad (3)$$

where N is the population size of a deme (assumed to be constant here), and m is the fraction of each deme which emigrates and is replaced by migrants drawn at random from the entire population. Other variations of the simple island model include stepping-stone models, in which populations

demes are connected to nearest neighbors on a lattice or ring structure (e.g. Kimura and Crow 1964; Kimura and Weiss 1964; Weiss and Kimura 1965; Kimura et al. 1968, 1971; Maruyama and Kimura 1980). While Model A is not a realistic model and given uniform information flow across demes, would not yield the differentiation into seriation solution groups that we see in archaeological data, it has a variety of known analytic results, making it useful for verifying the correctness of a computational model (see Section 5.2.2).

Model B depicts a snapshot in a more realistic metapopulation model, of the kind relevant in this research. The model either represents a synchronic view, or a snapshot in time of a metapopulation where demes can become extinct and be colonized. In this model, information flow between demes is variable among pairs, with very high flow rates between the two demes on the right, and very little flow between the bottom left deme and the other three. Model B is the type of metapopulation model that I believe best models what we see in regional-scale seriation solution groups given by Lipo's iterative pairwise method. I propose that the ability to seriate certain assemblages together (e.g., Group 2 in Figure A.3) is the result of assemblages representing samples of artifact discard from demes linked by high rates of information flow, with low rates of flow to other demes which will not seriate together, given a model of structured interaction like that in Model B.

The dynamics of neutral transmission within demes is the same as described in Section 3.2.1 above, with the exception that new variants are introduced to a deme both by endogenous innovation (μ_i), and by the probabilities of obtaining a variant from another specific deme (generalizing the notion of "migration" to not require residential mobility, but to include temporary traveling for trade) at rate $m_{i,j}$. Thus, for any given deme, the overall rate of innovation is:

$$\theta_i = 2N_i\mu_i + \sum_{j=1}^D m_j \quad (4)$$

Since modeling the transmission behind seriation solution groups requires that demes come into existence and then become extinct as identifiable units with population continuity, the set of demes available for migration will change over time, and necessarily so will the set of migration rate values (even if we do not model short-term fluctuations in migration rate within a pair of demes).¹⁰ Thus, the set of $m_{i,j}(t)$ values form a "migration matrix" $\mathbf{M}(t)$ which defines the metapopulation structure at any point in time, and the overall time-dependent migration matrix describes the history of population interaction.

It is this time-dependent migration matrix that we wish to model and understand, since I am proposing that its structure causes the overall pattern of archaeological assemblages being divided into sets of seriation solutions. The likelihood that information will be shared across two or more demes (and thus that they will have sets of archaeological class frequencies which are sufficiently "in sync" to seriate together) will be a function not just of the migration matrix possibilities just described, but the endogenous rate of innovation relative to the size of the deme (i.e., θ for each deme). The latter quantity controls the rate at which variants are lost to drift in a population. If two or more demes come to share a new innovation, then their migration matrix values over that time span must reflect one of the following scenarios: (a) demes partially overlap temporally, and have a relatively

¹⁰Since we are modeling unbiased transmission at regional scales here to explain variation among assemblages, and not reconstruct individual behavior, there is no need to model variation among individuals at all. This would change if we were studying selection at regional scales, of course.

high migration matrix term describing the sharing of variation between their source demes, (b) an assemblage represents the source of colonists for the establishment of a new deme, and thus there is one-way information flow between two otherwise non-overlapping demes, or (c) a combination of the two where a second deme is established by the first, and thereafter during a period of temporal overlap, the demes share migrants and information.

3.3. *Lipo's Iterative Pairwise Seriation Method*

Seriation is a method of ordering empirical entities using analytic classes, given an ordering principle for those analytic classes (Dunnell 1970). The most common use of seriation is to derive chronological orderings, so the classes employed are stylistic in character (e.g., "culture-historical types), and the ordering principles are heritable continuity of form and historical continuity. In other words, homology due to cultural transmission (Dunnell 1978; Lyman et al. 1997; O'Brien and Lyman 1999; Lyman and O'Brien 2006).

Ford's deterministic frequency seriation method has generally been non-quantitative, and was often done "by eye." This offers no means of evaluating the statistical fit of a solution to the ordering principles. Matrix methods, on the other hand, are inherently quantitative and could offer significance values for solutions, but collapsing the details of type frequencies into a single assemblage-level similarity value removes the possibility of determining the exactness of fit to unimodality for each type in a seriation. Thus, a hybrid approach is required. In his dissertation, Lipo (2001) elaborated a statistical approach to testing deterministic seriation orderings that grew out of work done with myself, R.C. Dunnell, and Tim Hunt (1997).

The first step is to incorporate the effects of sampling and sample size upon type frequencies; unimodality of class frequencies in a seriation can only be evaluated within confidence limits governed by the size of samples. Following Beals et al. (1945), a confidence interval is calculated for each type frequency in each assemblage, using the normal approximation for a confidence interval on a binomial variable (Cochran 2007).

Next, subsets of assemblages are found which seriate together (but cannot be further combined), within confidence limits. This step was facilitated by the use of an Excel-based macro for constructing and manually manipulating assemblages into groups.¹¹ Given a trial ordering of assemblages into groups which meet the distributional requirements for seriation, Lipo (2001) created a pairwise significance test to determine the likelihood that an entire ordering would occur simply due to chance sampling. One approach would be to calculate pairwise Student's t tests for type frequencies, but given the closed array of frequencies, repeated tests would rapidly lose statistical power, and there is no warrant to suppose that the sampling distribution of frequency differences in a seriation order are normally distributed. In fact, if we consider that an assemblage is a time averaged draw from the Ewens sampling distribution under the assumption of neutral transmission, we have every reason to believe that assemblage frequency differences are not normally distributed.

Thus, Lipo employed a Monte Carlo resampling approach to calculating a significance value. The process is outlined with a single pair of assemblages in Figure A.6. First, pairwise differences in frequencies are reduced to a series of directional comparisons (e.g., frequency of type 1 in assemblage A is less than the frequency of type 1 in assemblage B). Then, using the frequencies of types as the

¹¹The software is open-source, available under a Creative Commons license at <http://lipolab.org/seriation.html>. I am attempting to further automate this process, but if unsuccessful I will use the existing Excel solution in my dissertation research.

resampling distribution, random assemblages of the same size as the original assemblage are generated, and their type frequencies calculated. For each resampled assemblage, the directionality of type frequencies is tabulated. If a resampled assemblage has the same directionality of type frequencies as the original, a “match” is scored (see subfigure (b) in Figure A.6). If a resampled assemblage has different directionality for one or more types, that resample does not score a match. The significance value for the pairwise ordering is thus the proportion of matches seen. Lipo used 1000 resamples of each assemblage to perform pairwise comparisons, given that the method is computationally intensive (or was, in the late 1990’s). The end result is the ability to examine the goodness of fit of any trial seriation to the unimodal model expected from deterministic frequency seriation.

This method constructs seriation solution groups which meet the ordering principles (and thus measure heritable continuity and the flow of cultural traits through a set of interacting populations over time), and carry information about the precision and significance level of the ordering. Hereafter, when I use the term “seriation,” it will refer to this method, rather than the general universe of seriation techniques used by archaeologists.

4. Research Problems

In this section, I describe five specific research questions that I will address for my dissertation research. The first four questions primarily address issues concerning the methods needed to model unbiased cultural transmission at the level of a regional collection of artifact assemblages described by stylistic, paradigmatic classes, and observed through iterative frequency seriation. The simulation activities and analysis necessary to answer the first four questions should yield a regional transmission model capable of use in my case study of ceramic evolution in the St. Francis and Memphis study areas; that case study comprises the final question.

Research Question 1 (Adequacy of Unbiased Cultural Transmission).

Even though individual copying behavior is heterogeneous, with individuals displaying biases of different strength and “direction,” these biases may cancel out when observed at the level of whole populations, particularly when averaged over time. To what extent can archaeologists ignore models of bias from “dual-inheritance” theory and simply use unbiased transmission models to explain archaeological phenomena?

If we consider a simulated population with conformism and pro-novelty bias, each with a normal distribution of strengths, to what extent do population level observables such as trait richness, diversity, and trait lifetime depart from unbiased expectations? Can we detect departures from a Ewens sampling distribution in a heterogeneous population using the Slatkin Exact test? How do these answers vary with temporally aggregated observations?

Research Question 2 (Behavior of Unbiased Transmission in Design Space).

How do the quantitative descriptions of expected richness and evenness change as we observe trait transmission through the analytic filter of multi-dimensional, paradigmatic classes?

The traits of neutral models and cultural transmission simulations are not archaeological classes, and we cannot test archaeological hypotheses framed in the frequencies of types or classes against model traits. We must observe the dynamics of trait transmission with analytic classifications constructed in the same manner as the archaeological classes we use for describing artifact assemblages.

Because classifications are arbitrary with respect to the actual information which was transmitted between individuals in the past, and are set up to answer contemporary questions at particular temporal and spatial scales, we need to be aware that quantitative variables such as richness will not be the same as their values in the original “trait” space. The Ewens sampling distribution (see Eqns. 1 and 2) allow us to predict the number of traits we will see in a sample of a given size (see Equation 5, in Section 5.2.1). What is the comparable distribution for samples of size n , given a design space with a specific number of dimensions and modes (and thus, a specific total number of classes). What is the relationship for the frequency spectrum in classes, rather than traits? These questions aim to render the core probability distributions of the neutral model dynamically sufficient for the archaeological case.

Research Question 3 (What Metapopulation Dynamics Cause Multiple Seriation Solutions?).

In a metapopulation model of unbiased transmission, with variable “migration” rates, and demic colonization and extinction, what migration models, and what ratios of migration rate and endogenous innovation lead to a group of assemblages seriating together, distinct from other groups of assemblages?

Lipo’s iterative pairwise frequency seriation method partitions a set of assemblages into groups, each of which fits the expectations of the deterministic frequency seriation model within statistical tolerance limits. We understand this partitioning at a qualitative level to arise from differences in the degree of interaction between local communities. This research question addresses how partitioning of seriation solutions arises mechanistically in a spatial version of neutral theory. What kinds of population structures lead to partitioning of seriation solutions, and how does population structure covary with the endogenous innovation rate (since innovation within a deme tends to counteract the force of information flow from other demes which is the cause of “synchronization” of class frequencies). Is there, for example, a “critical ratio” of migration rates to endogenous innovation on a particular migration structure, below which demes cease being able to seriate together?

Research Question 4 (What Drives Richness and Evenness Patterns in Seriation Solutions?).

Seriation solution groups seem to vary in overall richness (the number of classes represented across all assemblages in a solution group) and average evenness (calculated for each assemblage and then averaged). Do particular models of migration and population structure lead to specific profiles of richness and evenness within a solution group?

Since richness and evenness are driven primarily by innovation, whether endogenous or from “migration,” this question asks how these variables are related to the population structures studied in Question 3. Are particular combinations of richness and evenness associated with specific models of migration, for example, such as clustering, or isolation by distance effects?

Research Question 5 (Accounting for Differences in PFG Seriation Solutions).

Given modeled relationships between seriation solution groups and interaction structure, which specific models of unbiased transmission in a regional metapopulation best account for each solution group in the PFG case study?

In existing seriation solutions, we see difference both in the spatial scope (and number of assemblages) present in solutions, and the mean richness and evenness of class frequencies in those solution groups. Which models have the maximum likelihood of describing the migration matrix structure (i.e., interaction structure) and innovation rates which lead to the seriation descriptions for groups of PFG assemblages in the Lower Mississippi River valley study area?

5. Research Methods

5.1. Computational Modeling

In this section, I describe the construction and verification of a computational or simulation model designed to address the research questions outlined above. Following this, I describe simulation experiments designed to answer the first four research questions, before turning to analyses of my Mississippi River valley case study.

5.1.1. Model Construction

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.¹² 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.¹³ **TransmissionFramework** runs on any platform capable of supporting a Java 1.6+ runtime, with optional scripts requiring Ruby 1.9+.

At the current time, **TransmissionFramework** already implements the WF-IA model for single populations, as described in Section 3.2.1, and possesses framework code for implementing structured and metapopulation models with migration or transmission between demes. It also has implementations of conformist and anti-conformist (or “pro-novelty”) bias, and an implementation of the continuous-time unbiased Moran model, which describes copying with overlapping generations.

Several features need to be implemented for my dissertation research, however, in order to tackle the research questions outlined in the previous section. In Appendix A, I describe the features, and note the development status of each at the time of writing. Those already present in the **TransmissionFramework** codebase in skeletal or partial form are noted, as are those which remain to be designed and implemented. In the next two subsections, I describe how I currently intend to implement classification and metapopulations in the computational model, since the implementations must match our expectations from scientific models and classificatory practice in archaeology.

5.1.2. Simulating Archaeological Classification

Currently, **TransmissionFramework** like most cultural transmission simulations has a single representation of cultural variants, and all counts and frequencies tracked have “traits” as the unit of both

¹²Forward-time approaches are not necessarily equivalent to “agent-based models,” but ABM techniques are useful in implementing forward-time models.

¹³**TransmissionFramework** can be downloaded or the code examined at <http://github.com/mmadsen/TransmissionFramework>.

transmission and observation. The framework is flexible enough, however, to “observe” other units which are functions of traits. I will implement paradigmatic classification (with multiple hierarchical levels) given this capability.

In the abstract, a paradigmatic classification is a set of dimensions along which variation can occur, split into modes or attributes which describe discrete portions of the variability in that dimension (Dunnell 1971). The classification itself is constructed by intersecting each dimension, to form their combinations (Figure A.7). Formally, the set of classes is the discrete product space of the dimensions. In a real classification used by archaeologists, each dimension and all of the modes would possess detailed definitions called *significata*, with each class thus possessing a necessary and sufficient definition for membership. In the abstract formal version described here, since I want to understand the quantitative effects of observing trait transmission through a classificatory filter of given dimensionality and granularity, definitions with archaeological content for classes are omitted from the simulation model. A classification model comprises (a) A set of traits which form the actual transmitted information, and where copying and innovation occurs. (b) Some number of observational dimensions, each specified by some number of attributes, and (c) A mapping which describes how traits are assigned to dimensions and modes for observation .

The mapping described in the third step is thus the analog of the *significata* for classes. Part of my dissertation research will involve finding good ways of specifying this mapping. At this point I have an initial proposal:

- Traits belong to a TraitDimension, which collects all of the variants (which are potentially infinite in the WF-IA model) which can occur in a particular chunk of the physical world.
- Traits are represented by real numbers in the interval $[0, 1]$, providing a potentially infinite number of innovations in any area of design space.
- Innovations occur at the normal rate (θ) for the WF-IA model, and are implemented for a focal trait by selecting a random point within $[0, 1]$ which is unoccupied (additional decimal precision can always be used to ensure that an innovation is new and unique).
- ClassDimensions map to TraitDimensions, since problem orientation in archaeological analysis should often yield dimensions of variability that are constrained enough by the physics and chemistry of construction processes to be space-like and invariant.¹⁴
- ClassModes are arbitrary partitions of the unit interval for each ClassDimension, however, to reflect the arbitrary nature of analytic classification with respect to the culturally transmitted information we model.
- ObservationClasses are then the product space of intersecting ClassDimensions together, with each ObservationClass representing a combination of ClassModes from the ClassDimension set.

¹⁴This is one aspect of this protocol that I find potentially problematic. In analyzing stylistic variation, this may be a reasonable approach, but in functional contexts, the relationship of multiple sets of ClassDimensions to TraitDimensions may be important to analyze, since we often cannot predict which sets of space-like engineering relationships constrained variation in artifact form. I am open to the notion that this will need to be done for stylistic decoration, but will begin by assuming a simpler link between the two.

- Hierarchical classification, if desired, is accomplished by grouping lower-level ClassModes together into sets, and amalgamating any counts or frequencies they store.

This scheme, although abstract, should fulfill the intent of separating unobservable cultural information which is the subject of actual transmission, from the observational units we construct to measure it. This scheme allows us to measure the effect of changing the “level” of classification (i.e., the number of dimensions and granularity of modes), as well as examining the effect of different ways of slicing up a given dimension upon our statistical analyses (by choosing different non-overlapping partitions of the unit interval comprising a ClassDimension).

This is depicted in a schematic graphical form in Figure A.8, to demonstrate how this scheme interacts with a new innovation. A dimension of transmitted variability (i.e., the unobservable cultural traits which were actually passed between individuals in the past) is depicted as the unit interval $[0, 1]$, with existing traits as small arrows occupying points on the line. Two archaeological dimensions (for simplicity of depiction, other dimensions in a classification have been suppressed) are shown, with one dimension nested hierarchically inside a second, providing two levels of resolution. For each dimension, mode boundaries are shown, created normally by the mode definitions. Blue numerals indicate the counts of traits which map to the modes in each dimension. In the bottom two blocks, I show alternative scenarios for an innovation occurring on top of the state shown above.

In the bottom two blocks, an innovation event occurs, and a new trait is created some distance away from the original trait held by an individual, with the individual then carrying the trait. There are two possible outcomes for this innovation event, since the mapping of the new trait to classificatory modes is arbitrary with respect to the random selection of location for the new trait.

In Time 2A, the innovation places it in a different mode, thus incrementing the count for the adjacent mode in both levels of the dimension hierarchy, and decrementing counts for the existing mode. In Time 2B, the innovation remains within the same mode, and the existing mode’s counts stay unchanged. Since our classifications are chosen with dimensionality and mode “granularity” which is arbitrary with respect to the innovation process being modeled, some innovations may be visible as changes in counts of archaeological classes, and others will not. The innovation shown in the Time 2A scenario alters archaeological class counts; the innovation shown in Time 2B does not.

This method of representing paradigmatic classification and separating evolving traits from their observational statistics also allows future experimentation with different models of innovation, particularly those involving perceptual and copying error (Eerkens and Lipo 2005). It would also allow easy representation of opinion dynamics models from economics and sociophysics, such as the Deffuant model whereby individuals with different traits “compromise” by choosing traits that are “closer together” with some probability (Deffuant et al. 2000; Fortunato 2004; Deffuant 2006; Stauffer and Meyer-Ortmanns 2004; Deffuant et al. 2005, 2002; Stauffer et al. 2004).¹⁵

5.1.3. Implementing Metapopulation Cultural Transmission

Implementing multiple population or metapopulation models in **TransmissionFramework** is mostly complete. Individual agents in the framework may be tagged with identifiers which allow any statistic to be measured for the subset of agents holding that identifier. An example might be “Deme 2” as a tag, in which case trait counts are calculated not only for the entire population, but separately for Deme

¹⁵In fact, my choice of the unit interval as the underlying trait space is derived from the Deffuant model.

2. Agents can have any number of tags, and tags can be changed. The latter feature allows easy implementation of permanent or temporary migration, by switching "Deme 2" for "Deme 6" and, possibly, back again.

What remains to be implemented is the modeling of a process for creating new demes and having existing demes go extinct. There are two requirements. First, it should be possible to specify a probability distribution governing deme lifetime and deme birth rate, and allow random configurations to unfold across many simulation runs. This allows the study and statistical analysis of transmission in evolving metapopulations in the abstract. Second, it should also be possible to provide a *configuration* for a metapopulation, with a specific number of demes, and an order of occupation and deme durations. This will allow representing particular empirical cases, such as assemblages from the PFG case study, so that we can analyze the properties of transmission across a specific region.

Finally, new demes should be populated by colonization from existing demes, and it should be possible to specify different models for this colonization. Slatkin (1977) distinguished between two extremes: a "migrant-pool" model whereby colonists for a new deme are drawn randomly from the whole metapopulation, and a "propagule pool" model, where colonists are chosen from a single extant deme in the population. The migrant-pool model will allow creation of island model scenarios useful for testing, while the propagule-pool model allows the modeling of more realistic empirical scenarios where settlements are derived from known sources, given historical continuity of artifact classes.

5.2. 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 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.

Computational models, whether they implement agent-based simulations, or Monte Carlo methods to solve systems of equations, are complex entities, subject to many sources of error. Models differ both from theory and from the real world in many respects. We often use simplifications of theory in order to make computations tractable. For example, my Research Question 1 essentially asks whether archaeologists need to use complex, heterogeneous models of transmission bias, or whether a simplified unbiased approximation is adequate given the typical spatiotemporal scale of analysis in archaeology. We often face difficulties in measuring model parameters. More fundamentally, computational models differ from reality because researchers select a subset of phenomena that form our research questions, and ignore many interactions and effects which are not of immediate interest.

Beyond science-based issues with computational models, the software which implements a computational model is often complex and composed of many modules, and often large amounts of external library code. It is thus important that scientific research employing a computational model be designed in such a way that the software codes be assessed, tested, and the results of such analysis documented. The extent to which simulations in archaeology are well-tested is unclear. For example, in a recent book (van der Leeuw and Kohler 2007), there are no entries for "testing" or "validation" in the index, and none of the chapters which employ computational models describe how the codes were

665 tested or assessed.¹⁶ Furthermore, archaeological publications employing simulation models rarely
666 describe the actual computational model in enough detail to understand its construction or function-
667 ing, although some authors make the code available upon request (while others do not, including some
668 of the most well-known models).

669 The problem of assessing the quality of computational models is important enough that the De-
670 partment of Energy and the Air Force Office of Scientific Research requested that the National Re-
671 search Council study the foundations of verification, validation, and uncertainty quantification (VVUQ)
672 activities for computational models in science and engineering. Their draft report forms the basis of
673 my approach to verification and uncertainty analysis in this research ([Committee on Mathematical](#)
674 [Foundations of Verification et al. 2012](#)).¹⁷

675 Verification answers the question, “how accurately does a computational model solve the underly-
676 ing equations of a theory for the observable quantities of interest.” In a more general sense, verification
677 addresses how well a computational system reflects the conceptual model an investigator has in mind,
678 but in this research I employ the NRC’s mathematically-oriented definition since it provides a clear
679 way to determine whether a model has been verified for the purposes at hand. A key recommenda-
680 tion from the NRC report is to employ a layered strategy to verify complex, multi-scale computational
681 models. In this research, I address verification at two scales: verification of the dynamics of unbiased
682 transmission within a single population, and verification of the dynamics within and between demes
683 in a metapopulation.

684 5.2.1. *Verification of Single-Population Unbiased Transmission*

685 **TransmissionFramework** is constructed to be easily testable, in order to verify its functionality as
686 new models are constructed or features added. Unit tests (or code verification) are a “best practice”
687 from software engineering that bundle integral software tests with the actual simulation code, to verify
688 that code acts as expected. An example from the current version of **TransmissionFramework** is a
689 test which automates the following sequence: (a) Construct a dimension with eight traits or modes;
690 (b) Construct 88 agents, and have each agent adopt one of the traits in different proportions; (c) Verify
691 that the total of trait counts across all traits at the end of all adoption events is 88. Other tests verify
692 different aspects of the code base, and taken together, unit tests provide a means of ensuring that the
693 low-level operations which compose a computational model are performing as specified.

694 Solution verification begins where unit tests end, and are often called “functional tests” in soft-
695 ware engineering, since they verify the proper functioning of a system from end to end: given known
696 inputs, does the system yield known good outputs? For unbiased transmission in a single popula-
697 tion, I verified a key observable quantity in the well-mixed Wright-Fisher infinite-alleles model within
698 **TransmissionFramework**.

699 The number of variants expected K_n in a sample of size n is a good test of the proper functioning
700 of a copying model because for a computational model to report a correct distribution of values for
701 K_n , both the copying rules and innovation rules in the model must be functioning correctly, and be
702 called in the correct proportions per unit time. [Ewens \(2004, Eq. 3.84\)](#) gives the full distribution of

¹⁶This is not to say that the book chapters are not valuable contributions to archaeology. I chose this example because it reports upon simulation projects which represent the state-of-the-art in contemporary archaeology.

¹⁷The NRC co-chair told me that the original charter for their analysis of computational modeling included the social sciences and especially economics, but that the committee could not find enough consistency or even examples of quality assessment to make a useful study (Adams, personal communication, April 2012).

703 K_n as:

$$\mathbb{P}(K_n = k) = \frac{|S_n^k| \theta^k}{\theta^{(n)}} \quad (5)$$

704 This distribution is difficult to use directly, so an approximation is usually required.¹⁸ Using the
 705 frequency spectrum of the WF-IA, [Ewens \(2004, Eq. 3.94\)](#) gives the expected value of K_n for a specific
 706 θ value:

$$\mathbb{E}(K_n) = \int_0^1 (1 - (1 - x)^n) \frac{\theta}{x} (1 - x)^{\theta-1} dx \quad (6)$$

707 I performed multiple simulation runs at θ values ranging from 2 to 40, for 5000 generations in
 708 a simulated population of 2000 individuals. Each parameter combination was represented by 3 sim-
 709 ulation runs. The initial transient behavior of the model is discarded from data analysis by skipping
 710 the first 750 generations, given the mixing time analysis by [Watkins \(2010\)](#). At each time step in a
 711 simulation run, the simulator took a sample of 30 individuals and tabulated the traits held by those
 712 individuals, and recorded the value of K_n . This yielded 408,478 samples of K_n across across validation
 713 runs. For each value of θ , I calculated the mean and standard deviation of K_n values.

714 Using Mathematica 8.0 with MathStatca 2.5 installed, I then calculated expected values for each
 715 θ level used in simulation, employing Equation (6). Table A.1 compares the expected values with the
 716 distribution of simulated values. In all cases, the analytical results are extremely close to the observed
 717 mean K_n values from simulation, and certainly well within 1 standard deviation. At least from the per-
 718 spective of K_n as a model observable, **TransmissionFramework** properly implements the well-mixed
 719 Wright-Fisher infinite-alleles model. Additional observable quantities may require verification during
 720 my dissertation research, and I will follow the same approach outlined here.

721 5.2.2. Verification of Metapopulation Transmission

722 Verifying the metapopulation implementation in **TransmissionFramework** builds upon the prior
 723 verification of unbiased transmission in a single population, as recommended by the [Committee on](#)
 724 [Mathematical Foundations of Verification et al. \(2012\)](#). I propose to verify the implementation using
 725 the simplest structure: the symmetric well-mixed island model (Figure A.5, Model A).

726 Equation 3 is Wright’s classic relationship describing the proportionality of trait diversity to migra-
 727 tion rates between populations in the island model. As [Whitlock and McCauley \(1999\)](#), describe, there
 728 are good reasons not to assume that F_{st} can be used to estimate migration rates in real populations.
 729 But for verification purposes, since we can construct a “perfect” island model with exactly equal mi-
 730 gration rates and known population sizes, estimation of F_{st} from samples taken from an island model
 731 simulation will be a good test of the correctness of both the migration code, and its interaction with
 732 the underlying Wright-Fisher infinite-alleles process occurring in each subpopulation. [Holsinger and](#)
 733 [Weir \(2009\)](#) and [Weir and Hill \(2002\)](#) describe several methods of estimating F-statistics from discrete
 734 genetic data, including the method of moments, which appears adequate for verification.

735 The approach will be the same as verification of the single-population simulation code. Simulation
 736 runs will be performed on a 10 population island model (a reasonable number of subpopulations

¹⁸MathStatca 2.5 running in Mathematica 8.0 is able to work with the distribution numerically, but cannot provide sym-
 bolic solutions for moments. Given the rising factorial, it is computationally expensive to calculate values from Eq. (5) for
 large values of θ and n .

helps ensure good estimates of variance between demes), with a range of migration rates, and a known innovation rate and population size per deme. Comparison of expected and observed F_{st} values will serve as the verification test.

It is also important to verify the diachronic aspects of deme creation and extinction. To do this easily, I propose to construct a simple metapopulation model with deme creation and extinction cycles in both **TransmissionFramework** and using **simuPOP** (Peng et al. 2012; Peng and Kimmel 2005), and compare multiple simulation runs with the same set of parameters. This type of “cross-validation” is not as strong as comparing code to analytic predictions, but if **TransmissionFramework** passes the previously described verification tests, as well as cross-validation with **simuPOP**, I will have strong confidence that the computational model correctly implements the models being studied in this research.

5.3. Model Fitting and Analytic Methods

Within archaeology, the fit of cultural transmission models to archaeological data is usually done through construction of a hypothesis test, comparing a test statistic to a theoretical distribution. Most frequently, the neutral or unbiased transmission model is employed as a “null model” or “null hypothesis,” and rejection of H_0 is considered evidence against neutrality and evidence for an alternative model, frequently involving biased transmission (e.g., conformism). The use of the null hypothesis testing (NHT) framework is standard in archaeological research, taught universally to students, and is reinforced by its use in statistical tests borrowed from theoretical population genetics, such as Slatkin (1994)’s neutrality test.¹⁹ Despite its ubiquity, NHT has several drawbacks for applications of cultural transmission models in archaeology.

First, we have no true “null models.” Although many authors have used neutral theory as a “null model,” it is not a true null hypothesis. The nature of a good null hypothesis is to describe the “no effect” or “by chance alone” case, against which we can detect a departure from such conditions through the determination of the tail probabilities of observed data. Researchers have tended to confuse the stochastic nature of “drift” in finite populations, and the unbiased nature of neutral cultural transmission with a hypothesis of “no effect.” Unbiased models of cultural transmission are, instead, complex models of a specific form of copying and diffusion within a population. The Wright-Fisher infinite-alleles model has different quantitative characteristics, for example, than a Wright-Fisher model which describes evolution among a fixed set of alleles (usually referred to as “k-alleles” models). Der et al. (2011) show that by modifying the third and higher moments of the Wright-Fisher process, the standard neutral model yields a variety of population processes which differ in their copying rules, and thus in observable predictions such as expected time to fixation or loss of alleles, without affecting the neutrality of alleles or drift given finite population size.

Archaeologists employing the “neutral model” have tended to ignore the great diversity which exists among models in mathematical population genetics, focusing only upon one expression of neutral evolution: the neutral well-mixed Wright-Fisher infinite-alleles model (described above), as brought into archaeology by Neiman (1995). Thus, even if we treated WF-IA as the our “standard” null model

¹⁹Hypothesis testing really comes in two forms: Neyman-Pearson theory, in which two substantive hypotheses are evaluated in a decision-theoretic procedure to select the hypothesis better supported by data, and Fisherian “significance testing,” in which a null hypothesis is examined for evidence of a significant departure from expectations. There are important differences between the two methodologies, to the point that Royall (1997) considers them separate statistical paradigms. For my purposes here, both types of procedure share the same limitations.

in cultural transmission studies, it is evident that its rejection in specific cases cannot be interpreted as strong evidence that our data do not result from other kinds of neutral or unbiased processes. Selecting and fitting an appropriate neutral model to an empirical data set is not as simple as failing to match a theoretical frequency distribution, as [Steele et al. \(2010\)](#) argue.

Second, too frequently we perform hypothesis tests on what amount to single data points, rather than sampling distributions of observations. Consider the use of Slatkin exact tests by [Steele et al. \(2010\)](#). Even though the two ceramic assemblages employed have good sample sizes ($n = 1392$ and $n = 2061$), the Slatkin test treats each assemblage as a single data point. The test determines the tail probability of an observed set of class counts, given the Ewens Sampling Distribution (Equation 2). If we set $\alpha = 0.05$ for our hypothesis test, we get a rejection region equal to the most extreme 0.025 percent of each tail. But a neutral WF-IA process will, occasionally, generate configurations that occur in this rejection region. This is easily verifiable by simulating WF-IA and tabulating the frequency with which Slatkin's test program returns a result $p < 0.025$ or $p > 0.0975$ ([Madsen 2012](#)). When we run a Slatkin test on a single assemblage, a rejection of the null hypothesis *could* mean that our sample was generated by a biased transmission process, or simply that it was one of the rare configurations of traits that are still possible under neutrality.

Granted, such Type I errors would only occur in our analyses rarely, but when they do, our interpretations of the test results will be incorrect. Can we do better within the hypothesis testing framework? Yes, by measuring the sampling distribution of Slatkin test results over many independent samples of class counts. If an overwhelming percentage of our test results are indicative of neutrality, that constitutes strong evidence that an unbiased transmission process was involved in generating our samples, even if a few samples fall into the rejection region. This approach is difficult in practice, since it either requires us to create sampling strategies during field recovery to recollect or resample areas, or it requires bootstrap sampling of a single large assemblage to examine neutrality tests on many subsets of the data.²⁰ The latter approach has not been employed to date by archaeologists interested in cultural transmission models.²¹

There are certainly ways to remedy the faults of the NHT approach for cultural transmission modeling, without abandoning the statistical methods commonly used by archaeologists. But in addition to being difficult and potentially expensive in many situations, NHT and hypothesis testing in general were designed for assessing the results of *experiments*, not uncontrolled observations. Ultimately, hypothesis testing is a highly useful set of statistical methods for situations in which a generating process is well understood, data collection well controlled and randomized, and where we can specify a “true null” of no effects, against we can test for departures. Such conditions rarely obtain in archaeology, and certainly not in the fitting and evaluation of complex cultural transmission models to real data.

5.3.1. *Strength of Evidence and Model Selection*

The alternative to hypothesis testing is an “evidentiary” or “strength of evidence” approach, in which the central task is to evaluate the degree to which data provide weak or strong evidence with regard to a suite of statistical models, each derived from a scientific (or “process”) model ([Anderson](#)

²⁰The latter strategy is obviously all one could do in analyzing previously published data, as much of the cultural transmission literature in archaeology has done. Even then, access to detailed rather than summary data is usually needed.

²¹I have not systematically explored whether this bootstrap approach will work well while retaining statistical power. **TransmissionFramework** is currently able to generate arbitrary Slatkin exact test samples from snapshots of the evolving population, so it would be relatively easy to test this conjecture. I have not proposed to do so here because it is tangential to my main research questions, but I may return to this in a conference paper.

2008; Royall 1997; Taper and Lele 2004). I propose to follow the procedures described by Burnham and Anderson (2002), based upon the information-theoretic approach advocated by (Edwards 1992; Royall 1997), and implementing Akaike’s modifications of the likelihood ratio test (Akaike 1973, 1974; Bozdogan 1987).²² In what follows I outline how AIC model selection will be used to address my research questions.

The first step is formulation of one or more scientific models we wish to assess as possible explanations for the variation seen in a data set. For example, in Research Question 2, the scientific model is the WF-IA model of unbiased copying, combined with a formal model of paradigmatic classification as employed by archaeologists. The research question asks about the statistical behavior of diversity measures for transmitted variation, when that variation is viewed through paradigmatic classes. Response variables would thus involve measures of class richness and evenness, calculated using counts of paradigmatic classes which are used to observe the underlying traits being transmitted in the simulation model.

The second step is then to formulate *statistical* models of the relationship between the response variables and scientific model parameters. In general, there should be multiple candidate models to describe this statistical relationship, because the point of model selection is to assess the strength of evidence our data provide for our statistical models, *relative to other possible models*. In order to evaluate these statistical models and select the best one given our data, we first fit each statistical model to the data derived from the analysis of the dynamics of our process model. In the above example, statistical models would attempt to infer the response variables from predictors such as the dimensionality of the classification and innovation rate of the transmission process. For each statistical model, maximum likelihood estimates (MLE) of statistical model parameters or regression coefficients need to be estimated ($\hat{\theta}_i$, where i refers to the i^{th} model in our statistical model set). Conceptually, the result of this step is an estimate of the likelihood of the parameter vector for each statistical model, given a data set: $\mathcal{L}(\hat{\theta}_i \mid \text{data}, \text{model}_i)$.

Given these estimates, we can then calculate the Akaike Information Criterion (AIC) for each model (Akaike 1973, 1974; Bozdogan 1987), given the parameter estimates and data. AIC is a wholly relative measure of the *expected information loss* in using each model to predict the data, given the MLE parameters. Despite being based upon the Kullback-Leibler divergence between a “true model” and an estimated model, AIC is not an absolute measure of model quality or fit, because we cannot know the “true model” in most circumstances. Instead, when we compare AIC values between models in a set, the unknown constant representing the true model’s information drops out and we are left with relative comparisons. AIC is composed of two terms. The first is the likelihood function of the model itself, given the data. The second term recognizes that more complex models, especially when nested, will always “improve” the fit of a model to specific data points, thus causing overfitting. Since we wish to minimize *generalization error*, and fit models which will be good at predicting many data sets, we seek to minimize fitting models to the noise or stochastic error components of our data. The second term in AIC (and most other model selection criteria) is thus a penalty term for the number of model parameters.

In practice, AICc is most frequently used, since it adds a correction to the penalty term for sample size, and thus is useful with small samples, and is asymptotically equivalent to AIC when sample size

²²Other model selection procedures exist, including k-fold cross-validation, but most model selection criteria have been shown to be asymptotically equivalent to AIC except in specific circumstances (Rao et al. 2001; Stone 1997).

is large (Hurvich and Tsai 1989). AICc is defined as:

$$AIC_c(i) = -2\log(\mathcal{L}(\hat{\theta}_i)) + 2K\left(\frac{n}{n - K - 1}\right) \quad (7)$$

for model i , where k is the number of model parameters, and n is the sample size for the dataset.

For the set of fitted models, the resulting set of AICc values can be compared (on an interval, not ratio scale) to make model selection decisions. A simple method is to find the minimum AICc value, and compute differences between this “best model” and the values for other models (Δ_i). The larger the Δ_i value, the less plausible it is that the model in question is a good model for the data as judged by its Kullback-Leibler divergence. Particularly for nested models, there are some simple rules of thumb, given by (Burnham and Anderson 2002). Δ_i values between 0 and 2 indicate substantial empirical support for a given model, while values greater than 10-12 indicate very little support.

The likelihood of a model, given the data, relative to other models, is given by:

$$\mathcal{L}(\text{model}_i \mid \text{data}) \propto \exp\left(-\frac{1}{2}\Delta_i\right) \quad (8)$$

We can then normalize the model weights to form model probabilities or “Akaike weights,” from which we can calculate the relative strength of evidence of two models.

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^R \exp\left(-\frac{1}{2}\Delta_r\right)} \quad (9)$$

where R is the number of statistical models in our comparison set. Conceptually, we have turned the likelihood of model parameters, given the data, into the likelihood of a model: $\mathcal{L}(\text{model}_i \mid \text{data}, \hat{\theta}_i)$.

The reason for the rules of thumb described above can be seen by examining the ratio of a model’s probability compared to the probability of the best model. For example, a model with a $\Delta_i = 4$ has a 7.4 to 1 likelihood compared to the best model. In other words, the best model is 7.4 times as likely to be a good empirical fit to the data as the model in question. Such evidence ratios decline swiftly with AICc differences. A model with $\Delta_i = 20$ has 22,000 to 1 odds of being the best model. In practice, we should give serious consideration to models with Δ_i values between 0 and 2, and even up to 4 if the data are particularly noisy. We should do so because even if our best model is 7.4 times more likely than the next best model, we can easily imagine collecting more or different data, and having this gap narrow or the ranking of models change. On the other hand, it is highly unlikely that a model 22,000 times worse than our current best model would improve its ranking no matter how much data we collect. We would thus discard models with values over 10 from further consideration.

Returning to the example of Research Question 2, the results of this process are a ranking of statistical models, describing the relationship between diversity measures and the key parameters of transmission models. This relationship is the answer to the research question. In some cases, it might be extremely difficult to decide between several models (say, if the Δ_i value is between 0-1). In analyzing empirical situations, we might want to keep a number of models, and use *model averaging* when we make predictions or inferences (Burnham and Anderson 2002). In situations where we are analyzing simulated data from theoretical models, this kind of outcome signals the need for additional modeling work to understand the best functional form for the relationship between variables.

Analysis for the datasets comprising my case study will proceed as above. The analysis will involve selecting the “best statistical models” for each of the process models I seek to distinguish in the case of LMV ceramic seriation solutions. The set of these best statistical models then becomes the candidate

model set for fitting to LMV seriation solution data (size distribution of solution clusters, richness and evenness within seriations). The process is repeated with parameter estimation and AICc calculation to answer the research question posed in my case study.

6. Simulation Experiments: Research Questions 1-4

The first four research questions lay the groundwork for employing iterative frequency seriation as an observational method for fitting regional-scale transmission models to archaeological data. Each of the questions also addresses a specific theoretical or methodological issue, using the computational model outlined above as a common toolkit. Each question has standalone value apart from their importance in my dissertation. I plan to prepare journal articles addressing research questions 1-4, and submit them to appropriate journals as the results from each question are available. Further, I ask my committee and chair for permission to embed those articles in the dissertation draft as chapters. These articles along with literature review, analysis and justification of methods and the computational model, and the case study from the PFG study area, will comprise the full dissertation draft.

6.1. Question 1: *Adequacy of Unbiased Cultural Transmission*

The first research question is wholly theoretical, and addresses whether archaeologists need to employ detailed models of individual-scale transmission rules in order to explain archaeological data, particularly at and above the assemblage scale. The alternative hypothesis, which I propose here, is that although we know individual humans display various transmission biases (e.g., [Boyd and Richerson 1985](#); [Henrich and Gil-White 2001](#); [Henrich 2001](#); [Henrich and Boyd 1998](#)), in a population heterogeneous for these rules, the statistical properties of social learning will converge at a population level to appear unbiased. This hypothesis is also a possible explanation for why Bentley's recent work on contemporary data sets (some of which I have collaborated upon), fit some of the expectations of random copying, even though we know modern individuals often display various transmission biases ([Hahn and Bentley 2003](#); [Herzog et al. 2004](#); [Bentley 2005, 2007](#); [Bentley et al. 2007a, 2009](#)).

I propose a simple experiment. `TransmissionFramework` already has implementations of conformist and pro-novelty bias, and agents can have heterogeneous copying (and innovation) rules. If the alternative hypothesis is correct, then heterogeneous populations which are mixtures of conformism and pro-novelty bias will display population-level outcomes for trait richness and evenness that fit the expectations of a comparable neutral model. The degree to which this is true will likely depend upon the relative strengths of conformist and pro-novelty bias, and the relative proportion of each strategy in the population. Thus, the desired outcome of this analysis is a "phase diagram" displaying the region of parameter space in which heterogeneous populations have unbiased population-level outcomes.

The results of this research question are relevant both to archaeologists and other social scientists who currently employ dual-inheritance theory or models of transmission bias. Thus, I will seek publication in a general purpose anthropology journal, or a journal widely read by those studying cultural transmission across several disciplines, such as *Journal of Theoretical Biology*.

6.2. Question 2: *Behavior of Unbiased Transmission in Design Space*

The second research question asks how the statistical properties of unbiased cultural transmission, and specifically the WF-IA neutral model, can be measured when trait evolution is observed through analytic classifications of the kind used in archaeology. Since the analyst controls the construction of a classification independent of the information which was transmitted, the statistical measures we

typically examine for transmission models, will necessarily be transformed into functions of the dimensionality and granularity of the classification.

I propose to address research question 2 by the following simulation protocol:

1. Construct a set of dimensions for observation, with at least five nested hierarchical levels. Few archaeological studies have employed more than two, or three at the extreme, but if our goal is to understand how the mean and variance of quantitative variables, such as class richness, scale with the dimensionality of a classification, a larger number of dimensions is necessary to understand whether scaling is linear or nonlinear, for example.
2. Construct several alternative sets of modes or attributes for each dimension, as arbitrary partitions of the unit interval. Each dimension should be cut into 2, 4, 8, and 16 modes, and for each level of mode granularity, at least eight different partition sets will be generated.
3. Each set of underlying parameters for the WF-IA model (population size, innovation rate) will be replicated 100 times for each configuration of the observational classification.
4. For each run, after the simulation reaches equilibrium (i.e., the number of “generations” specified by the mixing time analysis by [Watkins 2010](#)), samples of both traits and observational classes of size 50, 100, 500, and 1000 will be taken at intervals using the time-averaging protocol described in [Madsen \(2012\)](#).
5. Given the analysis in [Madsen \(2012\)](#), there seems to be two regions of behavior for the neutral WF-IA model: very small innovation rates, with $\theta < 1.0$ where copying dominates the dynamics of trait frequencies, and intermediate to high innovation rates, with $\theta > 1.0$ where innovation pressure dominates trait frequencies. A set of innovation rates will be chosen to cover these regions, but fewer values than employed in my recent analysis, where many of the θ levels showed little difference in dynamics.

Samples will be post-processed to describe the distribution of richness K_n and evenness values (using a normalized version of the t_f measure employed by [Neiman \(1995\)](#), called the “index of qualitative variation” or IQV ([Wilcox 1973](#)) and given by:

$$IQV = \left(\frac{k}{k-1}\right)\left(1 - \sum_{i=1}^k p_i^2\right) \quad (10)$$

Given these data, I will formulate a candidate set of statistical models with richness and evenness as measured by the simulated classifications, as response variables, and predictor variables including the size of the design space (i.e., number of classes, or simply dimensionality of the classification), and the parameters which drove the neutral WF-IA model (i.e., innovation rate). The best model will be chosen using the model selection approach described in detail in Section 5.3.

Although this is a stepping stone towards understanding the linkage between seriation solution groups (and how seriation solutions vary with respect to such statistics given different underlying transmission models), the results will also help us understand whether estimation of demographic history from cultural transmission models has a dynamically sufficient basis, or suffers from too much equifinality to be useful. I strongly suspect that it does not, and that estimates of θ , even in relative

terms, cannot tell us much about population history (a view which is supported by the work of Myers et al. 2008 from a different perspective), but demonstrating this definitively in the archaeological context will be a useful side-effect of the experiment.

Question 2 is purely of interest to archaeologists, and I plan to seek publication in a venue which regularly publishes articles on theory and method, such as the *Journal of Archaeological Method and Theory*.

6.3. Question 3: What Metapopulation Dynamics Cause Multiple Seriation Solutions?

The third research question asks how regional variation in population structure and information flow within a transmission process results in seriation solutions which are partitioned into sets, as occurs in the case study. The first step in analyzing this question is to constrain the notion of “population structure” and migration matrices to be studied, since there are an infinite set of possible structures one could study.

Several disciplines have studied “spreading” or diffusion processes on structures which approximate Figure A.5, Model B. Epidemiology and studies of information spreading on social networks have yielded a large body of literature on the effects of spatial structure on diffusion processes.²³ For my purposes, the results suggest that several factors matter in changing the ease with which information flows within a structured population:

1. The presence of absence of long-distance dispersal or connections;
2. The number of strong connections between subpopulations; alternatively, the “average degree” in a social network graph;
3. The presence of clustering among subpopulations, both spatially and in connections and migration flows.

Simple models of these factors are shown in Figure A.9, where “connections” indicate regular information flow and copying between at least some individuals in those demes. In Model 1, demes are connected to their spatial neighbors, but also possess long-distance links to demes which might not be neighbors. Model 2 is the opposite, with demes are connected only to spatial neighbors. Model 3 is the same as Model 2, but with much sparser connections among spatial neighbors. Model 4 introduces clustering of migration connections, such that there is constrained information flow among demes.

But these simple models are sufficient to begin addressing the research question at the level of resolution needed for regional-scale archaeological analysis, I believe. I propose to use these four models the basis for constructing simulated metapopulations, and performing seriations of the simulation output measured through analytic classification. The patterns of connections shown in these four models

²³The literature on this subject is now truly massive. Key references I consulted in preparing this proposal include, however: Barthélemy et al. (2005, 2004); Barthélemy et al. (2005); Andersson and Britton (2000); Boguñá et al. (2003); Castellano and Pastor-Satorras (2006); Colizza and Vespignani (2007); Keeling and Eames (2005); Keeling (2005); López-Pintado (2008); Andersson and Britton (2000); Ball et al. (1997); Boguñá et al. (2003); Boguñá and Pastor-Satorras (2002); Castellano and Pastor-Satorras (2006); Colizza and Vespignani (2007); Daley and Gani (1999); Eguíluz et al. (2003); Keeling and Eames (2005); Keeling (2005); López-Pintado (2008); May and Lloyd (2001); Moore and Newman (2000); Moreno et al. (2002); Newman (2002); Newman and Watts (1999); Ohtsuki and Nowak (2006); Pastor-Satorras and Vespignani (2001a,b); Rhodes and Anderson (1996); Sood and Redner (2005); Zanette and Risau-Gusmán (2008); Zanette and Gil (2006); Zanette (2002)

actually correspond to migration rates defined between demes in the metapopulation. The classification will be chosen to have the same dimensionality and structure as the PFG types which compose the case study. Additionally, although Figure A.9 displays a static snapshot of connections, simulations will employ a model where demes have durations and exit the model, with new demes being added with random links corresponding to the population structure model being tested. The diachronic aspect to population structure models is crucial for creating transmission simulations which will seriate like real archaeological data.

With respect to the four models, their structure leads to the following predictions. The first model, with long-distance connections, will not easily split into multiple seriation solutions without extremely high levels of endogenous innovation to “drift out” information flowing in from all parts of the metapopulation. In contrast, the fourth model with clustering should yield seriation solution groups for each of the densely connected structures, and the “boundaries” between seriation groups will correspond to the areas of sparser linkage. This will occur at intermediate and even low innovation rates.

At a given level of endogenous innovation, Model 2 should always yield larger seriation solution groups than Model 3, since lower migration and nearest-neighbor migration patterns should yield a classic “isolation by distance” dynamic. I propose to examine a range of endogenous innovation rates relative a fixed intensity of migration across populations configured as Model 2 and 3 to determine if a “critical ratio” of migration rate to endogenous innovation yields partitioning into seriation solutions, relative to the density of connections.

In real archaeological data sets, we sample a small fraction of the archaeological record, and thus the absolute number of assemblages in a seriation solution is less informative than the relative distribution of sizes of seriation solutions. Model selection will proceed by using the distribution of solution cluster sizes, and its variance in the metapopulation, as the response variables, and the design space size, statistical parameters of the network model describing the metapopulation, and transmission model parameters forming the predictor variable set. As above, model selection begins by formulating a set of candidate statistical models for correctly inferring the distribution of seriation solution sizes from the predictor variables, and then employing the information-theoretic methods described in Section 5.3.

There are very few studies of cultural transmission in an evolving population structure, and thus apart from the seriation and classification aspects, the general results of this research question could be of interest outside archaeology. I will first pursue a publication of the results for archaeologists, including a strong focus upon measuring migration and population structure given multiple seriation solution groups in a region. This type of article seems highly appropriate for the *Journal of Anthropological Archaeology*, given their history of publishing similar work (e.g., Lipo et al. 1997). At a later time, I may publish a short version of these results abstracted away from the archaeological methods involved, and submit it to *Journal of Theoretical Biology*, which publishes a great deal of interdisciplinary work on cultural transmission theory.

6.4. Question 4: What Drives Richness and Evenness Patterns in Seriation Solutions?

The fourth research question assumes the results of Question 3, and asks what factors of the transmission process in a metapopulation drive the differences in class richness and evenness seen in different seriation solution groups. Recall from Figures A.2 and A.3 that solution groups vary in the number of classes represented, and the degree to which a seriation solution is dominated by one or a few classes, or whether many classes have intermediate frequencies.

In the single-population neutral model, both richness and evenness are determined by the innovation rate. In a metapopulation, these variables will be driven both by innovation, but also the pattern and rates of migration between demes. For example, the larger a seriation solution group, the greater mean richness ought to be, at a constant rate of innovation. Since more demes are strongly connected, there are more demes for a rare trait or class to occupy and persist within the population, so larger seriation solution groups should have “reservoir effect,” with higher richness overall.

I propose to monitor richness and evenness patterns during the simulation runs executed to address Question 3, and examine the relationship between migration models and these variables. No additional simulation runs will need to be done, and these appear as separate research questions because they represent different aspects of examining the behavior of unbiased cultural transmission through seriation methods. The same analysis strategy will be used, with a set of candidate statistical models, with richness and evenness distributions across and within seriation solution groups as response variables, and the same predictor variables as described in the previous section. In publication, these results will be combined with the results from Question 3.

7. Model Inference for PFG Ceramic Assemblages

The goal of my case study is to apply the results of examining the four theoretical and methodological questions to explaining the quantitative characteristics of seriation solution groups from the Phillips, Ford and Griffin survey. Not only do seriation solutions, such as those depicted in Figure A.2 and A.3 indicate potential chronological relationships, but as proposed here, they are an observable, empirically sufficient window into the dynamics of cultural transmission in each geographic area during the Late Prehistoric period in the Mississippi River valley. In this section I briefly outline how I intend to apply the results obtained here to examine interaction and heritable variation in this region.

7.1. Assemblages and Preliminary Analysis

As described in Section 2, of the 383 assemblages collected in the original Phillips et al. (1951) survey, only 220 assemblages have decorated sherds. Of the assemblages with decorated sherds, only 63 assemblages have more than 50 decorated sherds (Table A.2). Six of these assemblages were also recollected by Lipo (2001) for his dissertation work, increasing the sample sizes at those localities.

For those assemblages not included in Lipo’s work, I will begin by examining sample size adequacy, using bootstrap examination of expected richness and its variance, following his bootstrapping method (Lipo 2001, p.43). I will include those assemblages with stable expected values for richness in further analysis. These assemblages will then be seriated using the iterative pairwise protocol described in Section 3.3 into maximal solution groups meeting the unimodality assumption.

7.2. Interaction Analysis

Given a set of seriation solutions from the selected PFG assemblages, I propose to test which of the population structure models (e.g., Figure A.9) best account for the groupings of assemblages across the study area. The variables used to describe seriation solutions are those studied above: class richness, evenness, and the “size” or scope of seriation groups in a relative sense across the region. These variables are measured relative to the modified PFG classification used by Lipo (2001) to conduct seriations (and employed here as well).

There is no simple way to determine a “fit” between population structure models and these quantitative descriptions of a seriation solution. I propose to use simulation once again to understand

the likelihood that a specific interaction structure leads to seriation solutions with the specified combination of descriptors, in the context of a given classification “design space.” The approach is an application of the model selection and multimodel inference approach described by (Burnham and Anderson 2002).

I propose the following approach:

- **TransmissionFramework** will be configured to operate with observational classifications having the same dimensionality and number of modes as the classes used to seriate PFG assemblages.
- The seriated assemblages will be embedded into each of the four population structure types studied above. Each real assemblage will represent a deme in the metapopulation, and will be surrounded by other demes. The real assemblages thus represent a sample of a larger regional population of localities participating in transmission.
- The exact number of these “synthetic” demes is not crucial, since their number will be constant across all simulation runs. Similarly, the rule for new deme colonization and existing deme extinction will be constant across runs. What is important are the *relative* likelihood values across different models of population structure.
- Using the same values of θ and other parameters used for the theoretical studies above, I will then perform a set of simulation runs of unbiased transmission across the simulated populations, with replicates for every combination of interaction model and other parameters.
- Each run is observed through simulated class frequencies. Given a large sample of class frequencies from each of the “real” assemblages, a random sample of simulated sherds the same size as the corresponding PFG assemblage will be taken.
- These simulated assemblages will then be seriated. This process will be bootstrapped from the simulated class frequencies to produce a “distribution” of seriation solutions for each interaction model and innovation rate.
- AIC distance between the actual PFG seriation data and these simulated seriation distributions then serves as a model selection criterion for the interaction model appropriate for each of the PFG seriation solutions (Burnham and Anderson 2002).

Although my main interest in this analysis is constructing and testing methods for making cultural transmission models empirically sufficient using seriation methods, a concrete outcome of this analysis should be a “map” of regional differences in interaction models and relative differences in innovation rates, over the duration of seriation solutions for the Late Prehistoric period in the PFG study area.

8. Estimated Timeline

The software implementation described above will take 2-3 months of concerted effort, including basic unit testing, beyond what is already in the codebase. This estimate is based upon the time already spent implementing **TransmissionFramework** for use in my SAA paper at the 2012 meetings, and construction of a predecessor simulation framework, for collaboration with Alex Bentley and Carl Lipo (Bentley et al. 2009; Madsen et al. 2008). Detailed tasks for remaining features are outlined in Appendix A, along with notes on status and possible duration for each feature.

The verification studies described in Section 5.2 should take less than a month since a large parameter space is not required to verify behavior, and analysis of the results using R is not difficult. The simulation study for Research Question 1 should take approximately two or three weeks including processing and analysis, since the parameter space is of comparable size to the study done for Madsen (2012).

The computational work for Research Question 2 is much more significant, however, since the parameter space is much larger given the configurations of observational classifications and underlying model parameters. My plan is to parallelize the simulation runs across a set of Amazon EC2 virtual machine instances, and do large batches of simulations in parallel, and employ another set of EC2 instances to perform the raw data post-processing. Given previous simulation runs, I anticipate needing at least a solid month of simulation time for this problem, and 10-14 days of processing and analysis time. Two months is a safe estimate overall.

Questions 3 and 4 are addressed by the same set of simulations, with different data collected and different statistical analyses. That said, the parameter space given random metapopulation models, with replicates, is large. This study will probably take another two months total for both research questions.

Seriation solutions for the expanded data set should not take long, but a month seems reasonable. The simulation modeling of PFG assemblages should take another month, perhaps two depending upon the size of the simulated metapopulation. I am optimistically hoping to be able to write up the results in six to nine months, including short articles for submission. The total timeline is about 18 months, and if begun in summer 2012 with the software changes already in progress, could result in submission of a dissertation draft before the end of calendar year 2013, for graduation by Spring 2014. This estimate is obviously tentative, but I want a schedule to structure the work give myself milestones towards completion. I will update the schedule as necessary.

9. Summary

This research proposes to address the major reasons why cultural transmission models have largely been dynamically and empirically insufficient tools for explaining variability seen in archaeological assemblages. First, formal models from theoretical population genetics are invariably employed without modification, and archaeologists have applied statistical results from the distribution of modeled “traits” as if they represented the results of identifying artifacts to archaeological classifications. I propose a method for studying the statistical outcomes of cultural transmission models using paradigmatic classifications at variable levels of dimensionality and mode granularity, and where the statistical measures are relative to a specific “design space.” Second, synchronic or equilibrium model predictions are compared to expected values rather than distributions of values, and such measures are rarely studied for their diachronic behavior, which is essential given the accretional and time-averaged nature of the archaeological record. The latter problem is rectified by following up previous work which proposes that culture-historical methods such as seriation are the appropriate diachronic observational methods with which to test transmission hypotheses against archaeological data.

I propose construction and analysis of a computational model incorporating solutions to both problems, and demonstrate the utility of the results by an expansion of the analysis begun by myself and colleagues in Lipo et al. (1997), and greatly expanded by Lipo (2001). I break this analysis into five research questions. The first four address specific theoretical and methodological questions, the results of which will be submitted to journals for publication. My dissertation draft will include the

four manuscripts (or publications), a review of the theoretical models, a discussion of the state of cultural transmission modeling in archaeology, documentation of the computational model and tests of its validity, and the results of my case study of Late Prehistoric ceramic assemblages in the Lower Mississippi River Valley.

References

- Abramowitz, M. and Stegun, L. 1965. *Handbook of Mathematical Functions*. Dover, New York.
- Akaike, H. 1973. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60(2):255–265.
- Akaike, H. 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.
- Anderson, D. 2008. *Model based inference in the life sciences: a primer on evidence*. Springer Verlag.
- Andersson, H. and Britton, T. 2000. *Stochastic epidemic models and their statistical analysis*. Springer.
- Bailey, G. 1981. Concepts, time-scales and explanations in economic prehistory. *Economic Archaeology: towards an integration of ecological and social approaches*, pages 97–118.
- Bailey, G. 2007. Time perspectives, palimpsests and the archaeology of time. *Journal of Anthropological Archaeology*, 26(2):198–223.
- Ball, F., Mollison, D., and Scalia-Tomba, G. 1997. Epidemics with two levels of mixing. *Annals of Applied Probability*, 7(1):46–89.
- Barthélemy, M., Barrat, A., and Pastor-Satorras, R. 2005. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of Theoretical Biology*.
- Barthélemy, M., Barrat, A., Pastor-Satorras, R., and Vespignani, A. 2004. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.*, 92(17):178701.
- Barthélemy, M., Barrat, A., Pastor-Satorras, R., and Vespignani, A. 2005. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of Theoretical Biology*, 235(2):275 – 288.
- Barton, N. and Whitlock, M. 1997. The evolution of metapopulations. *Metapopulation biology: ecology, genetics, and evolution*, pages 183–210.
- Beals, R. L., Brainerd, G. W., and Smith, W. 1945. *Archaeological Studies in Northeast Arizona*, volume 44 of *University of California Publications in American Archaeology and Ethnology*. University of California Press, Berkeley and Los Angeles.
- Bentley, R. 2007. Fashion versus reason-then and now. *ANTIQUITY-oxford*, 81(314):1071.
- 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.

- 1197 Bentley, R., Lipo, C., Herzog, H., and Hahn, M. 2007a. Regular rates of popular culture change reflect
1198 random copying. *Evolution and Human Behavior*, 28(3):151–158.
- 1199 Bentley, R., Madsen, M., and Ormerod, P. 2009. Physical space and long-tail markets. *Physica A:
1200 Statistical Mechanics and its Applications*, 388(5):691–696.
- 1201 Bentley, R. A. 2005. Academic copying, archaeology and the english language. *Antiquity*,
1202 80(307):196–201.
- 1203 Bentley, R. A., Lipo, C. P., Herzog, H. A., and Hahn, M. W. 2007b. Regular rates of popular culture
1204 change reflect random copying. *Evolution and Human Behavior*, 28:151–158.
- 1205 Bentley, R. A. and Shennan, S. J. 2003. Cultural transmission and stochastic network growth. *American
1206 Antiquity*, 68(3):459–485.
- 1207 Bettinger, R. and Eerkens, J. 1999. Point typologies, cultural transmission, and the spread of bow-and-
1208 arrow technology in the prehistoric great basin. *American Antiquity*, 64(2):231–242.
- 1209 Binford, L. 1981. Behavioral archaeology and the “pompeii premise”. *Journal of Anthropological Re-
1210 search*, pages 195–208.
- 1211 Boguñá, M., Pastor-Satorras, R., and Vespignani, A. 2003. Absence of epidemic threshold in scale-free
1212 networks with degree correlations. *Phys. Rev. Lett.*, 90(2):028701.
- 1213 Boguñá, M. and Pastor-Satorras, R. 2002. Epidemic spreading in correlated complex networks. *Physical
1214 Review E*.
- 1215 Boyd, R. and Richerson, P. 1985. *Culture and the Evolutionary Process*. University of Chicago Press,
1216 Chicago.
- 1217 Boyd, R. and Richerson, P. 2008. Response to our critics. *Biol. Philos.*, 23:301–315.
- 1218 Bozdogan, H. 1987. Model selection and akaike’s information criterion (aic): The general theory and
1219 its analytical extensions. *Psychometrika*, 52(3):345–370.
- 1220 Burnham, K. and Anderson, D. 2002. *Model selection and multimodel inference: a practical information-
1221 theoretic approach*. Springer Verlag.
- 1222 Castellano, C. and Pastor-Satorras, R. 2006. Non-mean-field behavior of the contact process on scale-
1223 free networks. *Phys Rev Lett*.
- 1224 Cochran, W. 2007. *Sampling techniques*. Wiley-India.
- 1225 Colizza, V. and Vespignani, A. 2007. Epidemic modeling in metapopulation systems with heteroge-
1226 neous coupling pattern: theory and simulations. *arXiv*, q-bio.PE.
- 1227 Committee on Mathematical Foundations of Verification, V., on Mathematical Sciences, U. Q. B.,
1228 Their Applications, D. o. E., and Physical Sciences, N. R. C. 2012. *Assessing the Reliability of Complex
1229 Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quan-
1230 tification*. The National Academies Press.

- 1231 Daley, D. and Gani, J. 1999. *Epidemic Modeling*. Cambridge University Press.
- 1232 Deffuant, G. 2006. Comparing extremism propagation patterns in continuous opinion models. *Journal*
1233 *of Artificial Societies and Social Simulation*, 9(3).
- 1234 Deffuant, G., Amblard, F., Weisbuch, G., and Faure, T. 2002. How can extremism prevail? a study based
1235 on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 5(4).
- 1236 Deffuant, G., Huet, S., and Amblard, F. 2005. An individual-based model of innovation diffusion
1237 mixing social value and individual benefit¹. *American Journal of Sociology*, 110(4):1041–1069.
- 1238 Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. 2000. Mixing beliefs among interacting agents.
1239 *Advances in Complex Systems*, 3(4):87–98.
- 1240 Der, R., Epstein, C., and Plotkin, J. 2011. Generalized population models and the nature of genetic
1241 drift. *Theoretical Population Biology*, 80:80–99.
- 1242 Dunnell, R. 1970. Seriation method and its evaluation. *American Antiquity*, 35(3):305–319.
- 1243 Dunnell, R. 1985. Archaeological survey in the lower mississippi alluvial valley, 1940-1947: A land-
1244 mark study in american archaeology. *American Antiquity*, 50(2):297–300.
- 1245 Dunnell, R. C. 1971. *Systematics in prehistory*. Free Press, New York.
- 1246 Dunnell, R. C. 1978. Style and function: A fundamental dichotomy. *American Antiquity*,
1247 43(2):192–202.
- 1248 Durrett, R. 2008. *Probability models for DNA sequence evolution*. Springer, second edition edition.
- 1249 Edwards, A. 1992. Likelihood, expanded ed. *Baltimore: Johns Hopkins*.
- 1250 Eerkens, J. and Lipo, C. 2005. Cultural transmission, copying errors, and the generation of vari-
1251 ation in material culture and the archaeological record. *Journal of Anthropological Archaeology*,
1252 24(4):316–334.
- 1253 Eguíluz, V., Hernández-García, E., Piro, O., and Klemm, K. 2003. Effective dimensions and percolation
1254 in hierarchically structured scale-free networks. *Physical Review E*.
- 1255 Ewens, W. 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*,
1256 3(1):87–112.
- 1257 Ewens, W. J. 2004. *Mathematical Population Genetics, Volume 1: Theoretical Introduction*. Springer, 2nd
1258 edition.
- 1259 Fortunato, S. 2004. The krause-hegselmann consensus model with discrete opinions. *Arxiv preprint*
1260 *cond-mat/0403670*.
- 1261 Grayson, D. and Delpech, F. 1998. Changing diet breadth in the early upper paleolithic of southwestern
1262 france. *Journal of Archaeological Science*, 25:1119–1129.
- 1263 Hahn, M. W. and Bentley, R. A. 2003. Drift as a mechanism for cultural change: an example from baby
1264 names. *Proceedings of the Royal Society Biology Letters B*, 270:S120–S123.

- 1265 Hamilton, M. J. and Buchanan, B. 2009. The accumulation of stochastic copying errors causes drift in
1266 culturally transmitted technologies: Quantifying clovis evolutionary dynamics. *Journal of Anthro-*
1267 *pological Archaeology*, 28:55–69.
- 1268 Hanski, I. and Gilpin, M. 1997. *Metapopulation biology: ecology, genetics, and evolution*, volume 454.
1269 Academic Press San Diego.
- 1270 Henrich, J. 2001. Cultural transmission and the diffusion of innovations: Adoption dynamics in-
1271 dicate that biased cultural transmission is the predominate force in behavioral change. *American*
1272 *Anthropologist*, 103(4):992–1013.
- 1273 Henrich, J. and Boyd, R. 1998. The evolution of conformist transmission and between-group differ-
1274 ences. *Evolution and Human Behavior*, 19:215–242.
- 1275 Henrich, J. and Gil-White, F. 2001. The evolution of prestige: freely conferred deference as a
1276 mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*,
1277 22:165–196.
- 1278 Herzog, H. A., Bentley, R. A., and Hahn, M. W. 2004. Random drift and large shifts in popularity of
1279 dog breeds. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 271:S353–S356.
- 1280 Holsinger, K. and Weir, B. 2009. Genetics in geographically structured populations: defining, estimat-
1281 ing and interpreting fst. *Nature Reviews Genetics*, 10(9):639–650.
- 1282 Hurvich, C. and Tsai, C. 1989. Regression and time series model selection in small samples. *Biometrika*,
1283 76(2):297–307.
- 1284 Jordan, P. and Shennan, S. 2003. Cultural transmission, language, and basketry traditions amongst the
1285 california indians. *Journal of Anthropological Archaeology*, 22(1):42–74.
- 1286 Kandler, A. and Laland, K. 2009. An investigation of the relationship between innovation and cultural
1287 diversity. *Theoretical population biology*, 76(1):59–67.
- 1288 Kandler, A. and Steele, J. 2009. Innovation diffusion in time and space: effects of social information
1289 and of income inequality. *Diffusion Fundamentals*, 3:1–17.
- 1290 Keeling, M. 2005. The implications of network structure for epidemic dynamics. *Theoretical population*
1291 *biology*.
- 1292 Keeling, M. and Eames, K. 2005. Networks and epidemic models. *Journal of the Royal Society Interface*,
1293 2:295–307.
- 1294 Kimura, M. and Crow, J. 1964. The number of alleles that can be maintained in a finite population.
1295 *Genetics*, 49(4):725.
- 1296 Kimura, M. et al. 1968. Genetic variability maintained in a finite population due to mutational pro-
1297 duction of neutral and nearly neutral isoalleles. *Genet. Res.*, 11:247–269.
- 1298 Kimura, M., Maruyama, T., et al. 1971. Pattern of neutral polymorphism in a geographically structured
1299 population. *Genet. Res.*, 18(1):125–131.

- 1300 Kimura, M. and Weiss, G. 1964. The stepping stone model of population structure and the decrease
1301 of genetic correlation with distance. *Genetics*, 49(4):561.
- 1302 Kingman, J. 1977. The population structure associated with the ewens sampling formula. *Theoretical*
1303 *Population Biology*, 11(2):274–283.
- 1304 Kohler, T. A., VanBuskirk, S., and Ruscavage-Barz, S. 2004. Vessels and villages: evidence for con-
1305 formist transmission in early village aggregations on the pajarito plateau, new mexico. *Journal of*
1306 *Anthropological Archaeology*, 23(1):100–118.
- 1307 Landau, D. and Binder, K. 2005. *A guide to Monte Carlo simulations in statistical physics*. Cambridge
1308 Univ Pr.
- 1309 Lewontin, R. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press.
- 1310 Lipo, C. and Madsen, M. 2000. Neutrality, "style," and drift: Building methods for studying cultural
1311 transmission in the archaeological record. In Hurt, T. D. and Rakita, G. F. M., editors, *Style and*
1312 *Function: Conceptual Issues in Evolutionary Archaeology*, pages 91–118. Bergin and Garvey, Westport,
1313 Connecticut.
- 1314 Lipo, C., Madsen, M., Dunnell, R., and Hunt, T. 1997. Population structure, cultural transmission,
1315 and frequency seriation. *Journal of Anthropological Archaeology*, 16(4):33.
- 1316 Lipo, C. P. 2001. *Science, Style and the Study of Community Structure: An Example from the Central*
1317 *Mississippi River Valley*. British Archaeological Reports, International Series, no. 918, Oxford.
- 1318 López-Pintado, D. 2008. Diffusion in complex social networks. *Games and Economic Behavior*,
1319 62(2):573 – 590.
- 1320 Lyman, R. 2003. The influence of time averaging and space averaging on the application of foraging
1321 theory in zooarchaeology. *Journal of Archaeological Science*, 30(5):595–610.
- 1322 Lyman, R. and O'Brien, M. 2006. *Measuring time with artifacts*. University of Nebraska Press.
- 1323 Lyman, R., O'Brien, M., and Dunnell, R. 1997. *The rise and fall of culture history*. Springer.
- 1324 Madsen, M. E. 2012. Unbiased cultural transmission in time-averaged archaeological assemblages.
1325 *ArXiv e-prints*, 1204.2043.
- 1326 Madsen, M. E., Lipo, C. P., and Bentley, R. A. 2008. Explaining seriation patterns through network-
1327 structured cultural transmission models. In *Poster presented at the 73rd Annual Meeting of the Society*
1328 *for American Archaeology*.
- 1329 Maruyama, T. and Kimura, M. 1980. Genetic variability and effective population size when local
1330 extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy*
1331 *of Sciences*, 77(11):6710.
- 1332 May, R. and Lloyd, A. 2001. Infection dynamics on scale-free networks. *Physical Review E*.
- 1333 Moore, C. and Newman, M. 2000. Epidemics and percolation in small-world networks. *Physical Review*
1334 *E*.

- 1335 Moreno, Y., Pastor-Satorras, R., and Vespignani, A. 2002. Epidemic outbreaks in complex heteroge-
1336 neous networks. *The European Physical Journal B-Condensed Matter*
- 1337 Myers, S., Fefferman, C., and Patterson, N. 2008. Can one learn history from the allelic spectrum?
1338 *Theoretical population biology*, 73(3):342–348.
- 1339 Neiman, F. D. 1990. *An Evolutionary Approach to Archaeological Inference: Aspects of Architectural Vari-*
1340 *ation in the 17th Century Chesapeake*. PhD thesis, Yale University.
- 1341 Neiman, F. D. 1995. Stylistic variation in evolutionary perspective: Inferences from decorative diver-
1342 sity and interassemblage distance in illinois woodland ceramic assemblages. *American Antiquity*,
1343 60(1):7–36.
- 1344 Newman, M. 2002. Spread of epidemic disease on networks. *Physical Review E*, 66(1):16128.
- 1345 Newman, M. and Watts, D. 1999. Scaling and percolation in the small-world network model. *Physical*
1346 *Review E*.
- 1347 O’Brien, M. and Lyman, R. 1998. *James A. Ford and the growth of Americanist archaeology*. Univ of
1348 Missouri Pr.
- 1349 O’Brien, M. and Lyman, R. 1999. *Seriation, stratigraphy, and index fossils: The backbone of archaeological*
1350 *dating*. Plenum Pub Corp.
- 1351 O’Brien, M., Lyman, R., Glover, D., and Darwent, J. 2003. *Cladistics and archaeology*. University of
1352 Utah Press Salt Lake City, UT.
- 1353 O’Brien, M., Lyman, R., Mesoudi, A., and VanPool, T. 2010. Cultural traits as units of analysis. *Philo-*
1354 *sophical Transactions of the Royal Society B: Biological Sciences*, 365(1559):3797–3806.
- 1355 Ohtsuki, H. and Nowak, M. A. 2006. The replicator equation on graphs. *J Theor Biol*, 243(1):86–97.
- 1356 Pastor-Satorras, R. and Vespignani, A. 2001a. Epidemic dynamics and endemic states in complex
1357 networks. *Physical Review E*.
- 1358 Pastor-Satorras, R. and Vespignani, A. 2001b. Epidemic spreading in scale-free networks. *Phys. Rev.*
1359 *Lett.*, 86(14):3200–3203.
- 1360 Peng, B. and Kimmel, M. 2005. simupop: a forward-time population genetics simulation environment.
1361 *Bioinformatics*, 21(18):3686–3687.
- 1362 Peng, B., Kimmel, M., and Amos, C. 2012. *Forward-Time Population Genetics Simulations: Methods,*
1363 *Implementation, and Applications*. Wiley-Blackwell.
- 1364 Phillips, P., Ford, J., and Griffin, J. 1951. *Archaeological Survey in the Lower Mississippi Alluvial Valley,*
1365 *1940-1947*. Papers of the Peabody Museum of American Archaeology and Ethnology 25. Harvard
1366 University, Cambridge.
- 1367 Premo, L. and Scholnick, J. 2011. The spatial scale of social learning affects cultural diversity. *American*
1368 *antiquity*, 76(1):163–176.

- 1369 Prohaska, S. and Stadler, P. 2008. Genes. *Theory in Biosciences*, 127(3):215–221.
- 1370 Rao, C., Wu, Y., Konishi, S., and Mukerjee, R. 2001. On model selection. *Lecture Notes-Monograph*
1371 *Series*, pages 1–64.
- 1372 Rhodes, C. and Anderson, R. 1996. Dynamics in a lattice epidemic model. *Physics Letters A*.
- 1373 Royall, R. 1997. *Statistical evidence: a likelihood paradigm*, volume 71. Chapman & Hall/CRC.
- 1374 Schiffer, M. 1987. *Formation Processes of the Archaeological Record*. University of New Mexico Press,
1375 Albuquerque.
- 1376 Scholnick, J. 2010. *Apprenticeship, Cultural Transmission and the Evolution of Cultural Traditions in*
1377 *Historic New England Gravestones*. PhD thesis, University of Arizona.
- 1378 Shennan, S. 2011. Descent with modification and the archaeological record. *Philosophical Transactions*
1379 *of the Royal Society B: Biological Sciences*, 366(1567):1070–1079.
- 1380 Shennan, S. and Wilkinson, J. 2001. Ceramic style change and neutral evolution: A case study from
1381 neolithic europe. *American Antiquity*, 66(4):577–593.
- 1382 Slatkin, M. 1977. Gene flow and genetic drift in a species subject to frequent local extinctions. *Theo-*
1383 *retical population biology*, 12(3):253–262.
- 1384 Slatkin, M. 1994. An exact test for neutrality based on the ewens sampling distribution. *Genetical*
1385 *Research*, 64(01):71–74.
- 1386 Slatkin, M. 1996. A correction to the exact test based on the ewens sampling distribution. *Genetical*
1387 *research*, 68(03):259–260.
- 1388 Sood, V. and Redner, S. 2005. Voter model on heterogeneous graphs. *Phys Rev Lett*.
- 1389 Stauffer, D. and Meyer-Ortmanns, H. 2004. Simulation of consensus model of deffuant et al. on a
1390 barabasi-albert network. *International Journal of Modern Physics C*, 15:241–246.
- 1391 Stauffer, D., Sousa, A., and Schulz, C. 2004. Discretized opinion dynamics of the deffuant model on
1392 scale-free networks. *Journal of Artificial Societies and Social Simulation*, 7(3).
- 1393 Steele, J., Glatz, C., and Kandler, A. 2010. Ceramic diversity, random copying, and tests for selectivity
1394 in ceramic production. *Journal of Archaeological Science*, 37(6):1348–1358.
- 1395 Stein, J. 1987. Deposits for archaeologists. *Advances in archaeological method and theory*, 11:337–395.
- 1396 Stone, M. 1977. An asymptotic equivalence of choice of model by cross-validation and akaike’s crite-
1397 rion. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 44–47.
- 1398 Taper, M. and Lele, S., editors 2004. *The nature of scientific evidence: statistical, philosophical, and em-*
1399 *pirical considerations*. University of Chicago Press Chicago:.
- 1400 van der Leeuw, S. and Kohler, T. 2007. The model-based archaeology of socio-natural systems. *School*
1401 *of Advanced Research, Santa Fe*.

- Wakeley, J. 2008. *Coalescent Theory*. Harvard University Press.
- Watkins, J. 2010. Convergence time to the ewens sampling formula. *Journal of Mathematical Biology*, 60:189–206.
- Weir, B. and Hill, W. 2002. Estimating f-statistics. *Annual Review of Genetics*, 36(1):721–750.
- Weiss, G. and Kimura, M. 1965. A mathematical analysis of the stepping stone model of genetic correlation. *Journal of Applied Probability*, pages 129–149.
- Whitlock, M. and McCauley, D. 1999. Indirect measures of gene flow and migration: $F_{ST} \approx 1/(4nm + 1)$. *Heredity*, 82(2):117–125.
- Wilcox, A. 1973. Indices of qualitative variation and political measurement. *The Western Political Quarterly*, 26(2):325–343.
- Wright, S. 1943. Isolation by distance. *Genetics*, 28:114–137.
- Zanette, D. 2002. Dynamics of rumor propagation on small-world networks. *Physical Review E*.
- Zanette, D. and Gil, S. 2006. Opinion spreading and agent segregation on evolving networks. *Physica D: Nonlinear Phenomena*.
- Zanette, D. and Risau-Gusmán, S. 2008. Infection spreading in a population with evolving contacts. *Journal of Biological Physics*.

1418 Appendix A. Software Development Plan

1419 The following tasks are necessary in order to use **TransmissionFramework** to address the research
1420 questions proposed here. The status of each task is current as of 2012-08-30.

1421 Separate transmission traits and observational units

1422 Traits and trait dimensions that comprise the information flowing within a population should be
1423 separate from a set of observational units, which can be mapped onto traits and trait dimensions
1424 in arbitrary ways. Traits should continue to be trackable by count or frequency, to allow us
1425 to understand the effect of observing trait dynamics through observational classes. [STATUS:
1426 *partial* TIMELINE: *In progress summer 2012, requires 2 weeks solid effort*]

1427 Paradigmatic classification

1428 Simulation models should be configurable with one or more paradigmatic classifications, each
1429 of which specifies a number of dimensions and for each dimension, a number of modes. Di-
1430 mensions are attached to Trait Dimensions, and it should be possible to manually specify the
1431 mapping of modes to segments of a Trait Dimension, or to generate random partitions of a Trait
1432 Dimension into modes. [STATUS: *partial* TIMELINE: *In progress summer 2012, requires 1 week*
1433 *of solid effort on top of observational units infrastructure*]

1434 Individuals assignable to deme or local sub-population [STATUS: *complete* TIMELINE: *complete*]

1435 Deme-level frequency counting

1436 Count/frequencies of observational units must be tracked for the metapopulation as a whole
1437 and for demes. [STATUS: *partial* TIMELINE: *On hold, mainly needs testing after classification*
1438 *features are done, possibly 2 days of work*]

1439 Deme creation and destruction

1440 Because archaeological samples are not fully contemporaneous and may not overlap in time, the
1441 model should allow new demes to enter the simulation, and for demes to go extinct. Since this
1442 uses the existing tagging mechanism, mostly this is wiring and testing. [STATUS: *unimplemented*
1443 TIMELINE: *Autumn 2012*]

1444 Population of demes

1445 New demes should be populated either by colonization from a single “parent” deme, or by sam-
1446 pling individuals from the entire population (giving the classic island model). [STATUS: *unim-*
1447 *plemented* TIMELINE: *Autumn 2012*]

1448 Configuration of population and deme profiles

1449 Simulation models need a “profile” of what demes and population to create at run start, how
1450 demes and population evolve over time. This needs to be configurable to test randomized pro-
1451 files, as well as generate data in configurations that match the parameters of empirical examples.
1452 [STATUS: *unimplemented* TIMELINE: *Autumn 2012*]

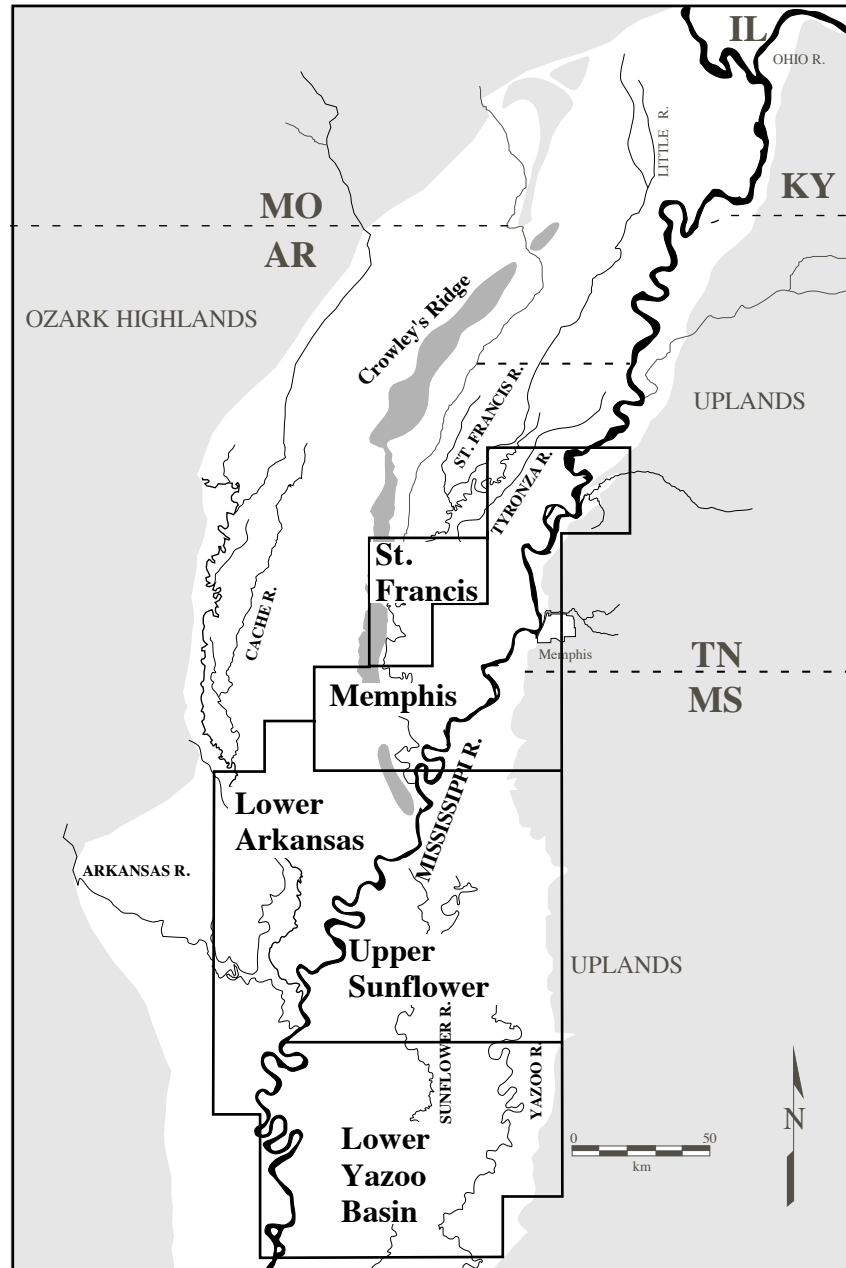


Figure A.1: Subdivision of the [Phillips et al. \(1951\)](#) study area into arbitrary analytic units by James Ford. Reproduced with permission from [Lipo \(2001, Figure 2.3\)](#).

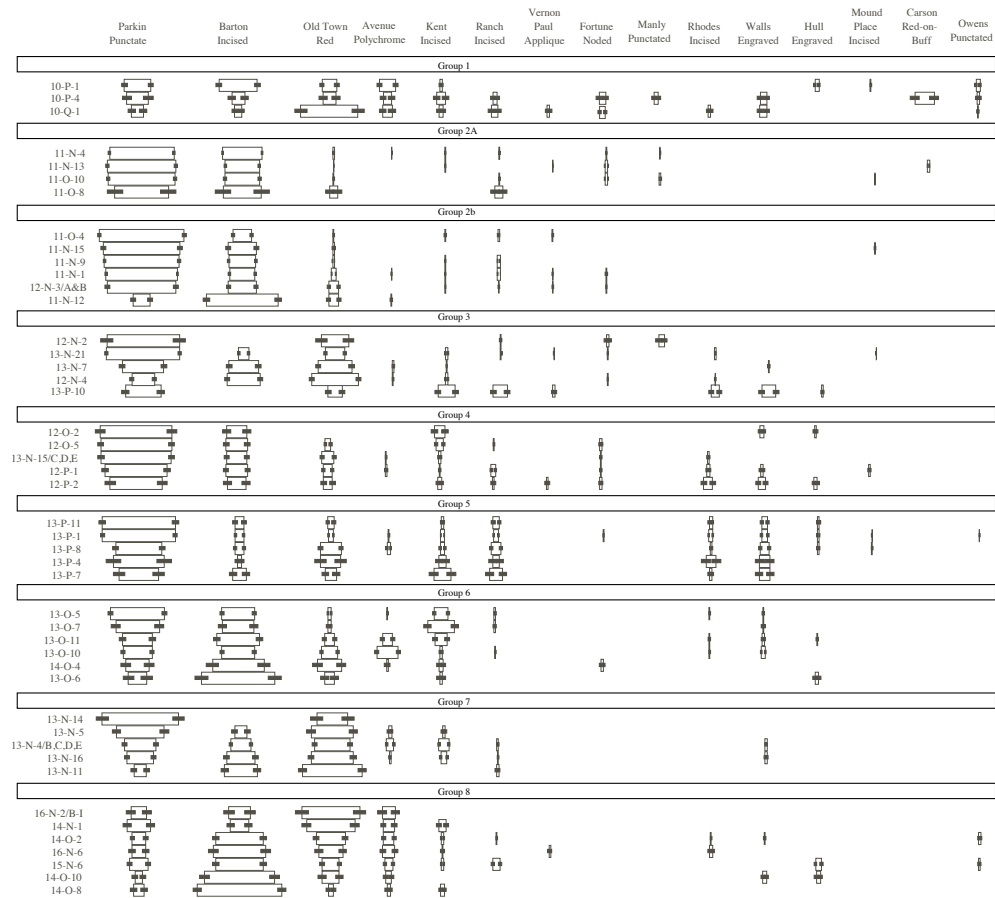


Figure A.2: Results of iterative deterministic seriation by Carl Lipo for PFG assemblages in the combined Memphis and St. Francis area. The eight solution groups are the largest seriations that could be formed using the assemblages without violating the requirement of unimodality within binomial error limits. Reproduced with permission from [Lipo \(2001, Figure 4.4\)](#).

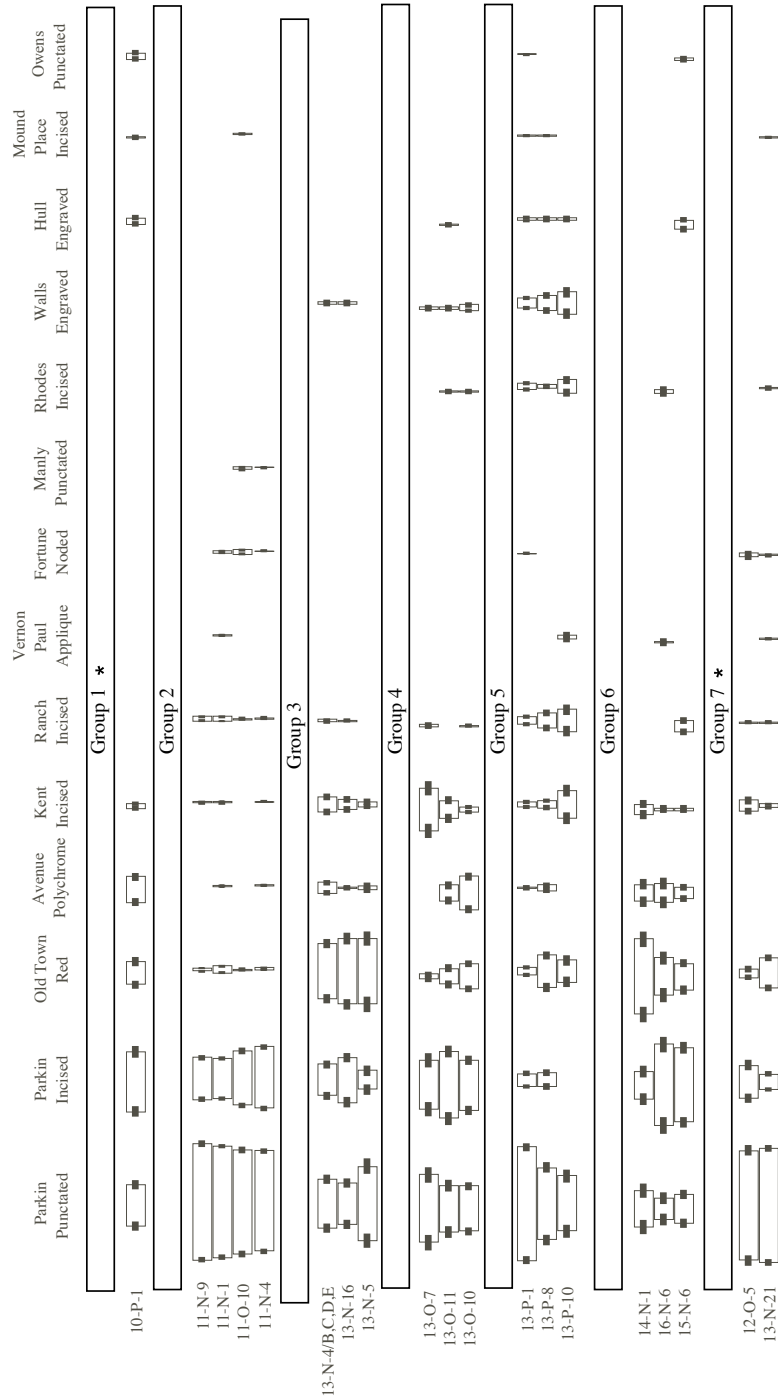


Figure A.3: Results of iterative seriation with pairwise significance testing. Groups 1 and 7 have no implied order given < 3 assemblages. Reproduced with permission from Lipo (2001, Figure 4.11).

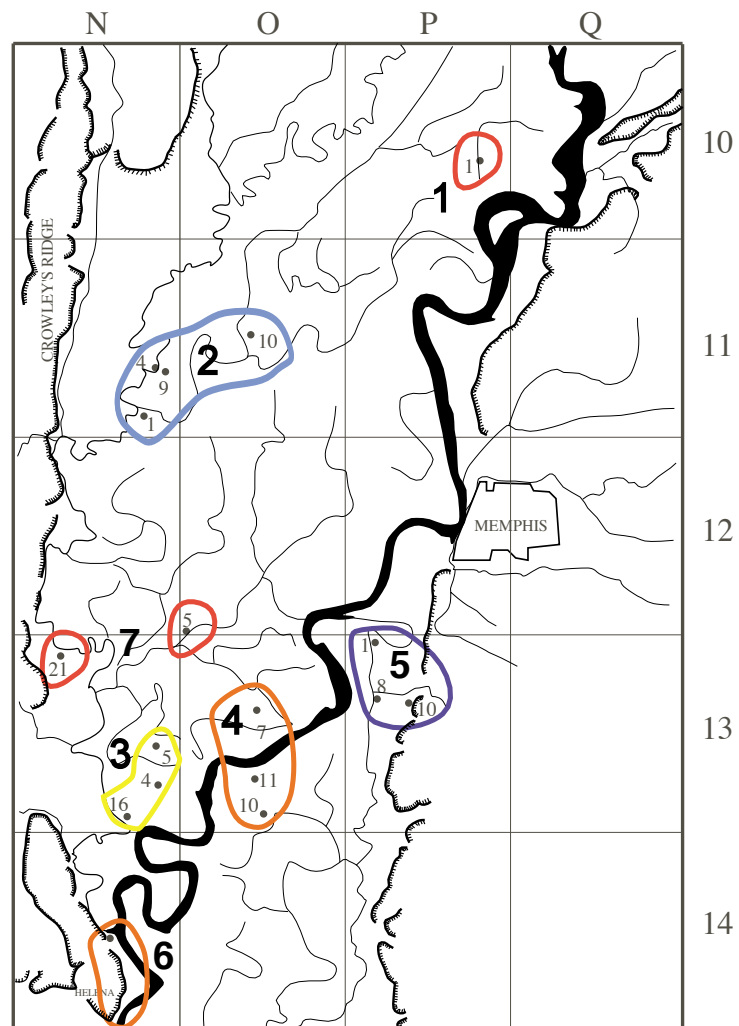
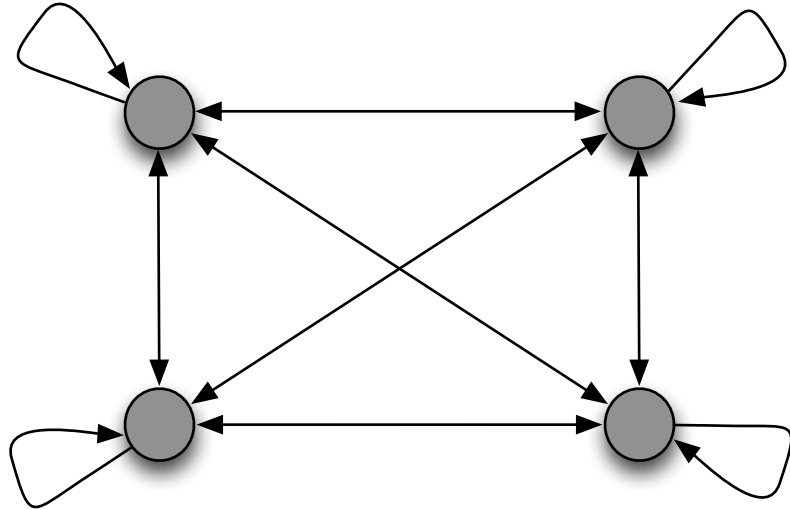


Figure A.4: Spatial pattern of assemblages from seriation solutions in Figure A.3. Reproduced with permission from Lipo (2001, Figure 4.10).

(a)



(b)

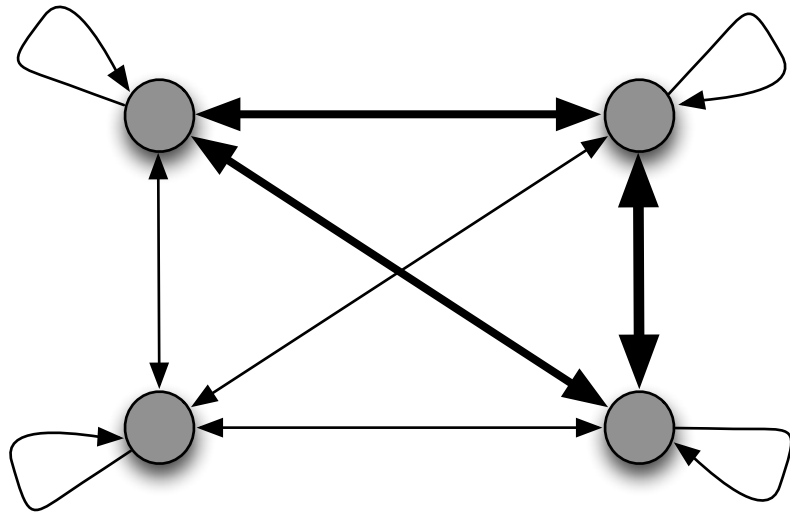


Figure A.5: Examples of simple metapopulation models. Each circle represents a deme or local population, and arrows are unbiased transmission flows, either within the deme itself (represented by the looped arrow), or between demes. Model (a) depicts the well-mixed version of a metapopulation model, with homogeneous flows between all demes; this model is equivalent to Wright's classical "island" model in population genetics. Model (b) depicts a metapopulation model with variable flows between demes, with arrow thickness indicating the rate of information flow.

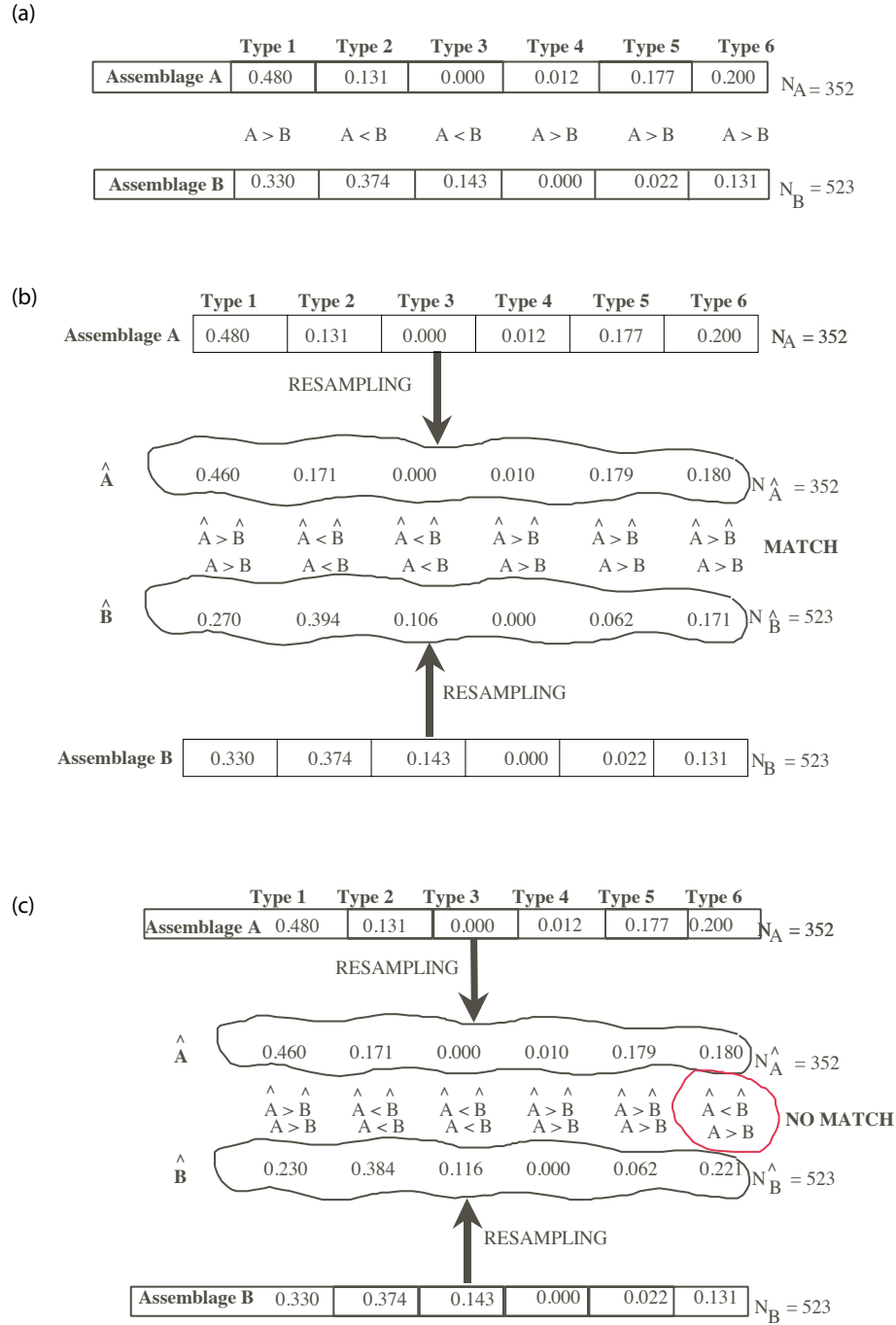


Figure A.6: Iterative pairwise significance testing for frequency seriation (Lipo 2001). (a) Counts are converted to frequencies, and the “directionality” of frequency comparisons is calculated for each type. Given the frequencies of types, bootstrap sampling is used to draw a set of random assemblages. For each random assemblage draw, the directionality of each type frequency is evaluated. If all types in a random assemblage have the same directionality, a match is scored, as in (b). If a resampled assemblage has different directionality for one or more types, no match is scored. The proportion of matches in bootstrapped assemblages constitutes the p value for the test. Redrawn with permission from Figures 3.9 - 3.11 in Lipo (2001).

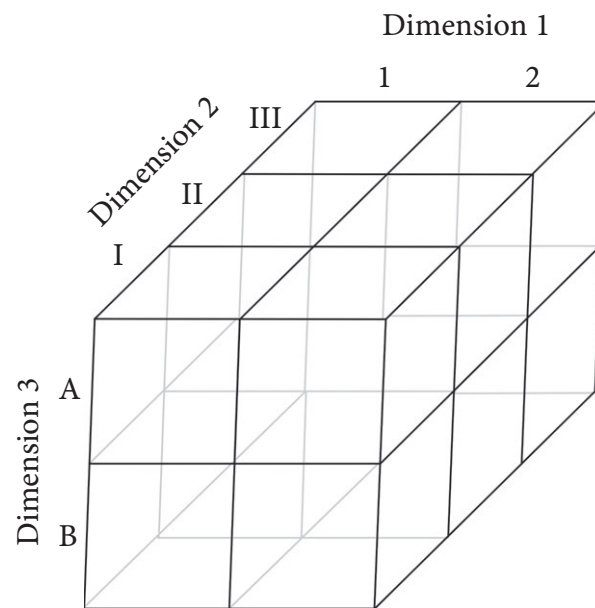


Figure A.7: Schematic example of a paradigmatic classification. This classification employs three dimensions, two of which are described by two modes apiece, the third being described by three modes. The result is a design space composed of twelve mutually-exclusive classes.

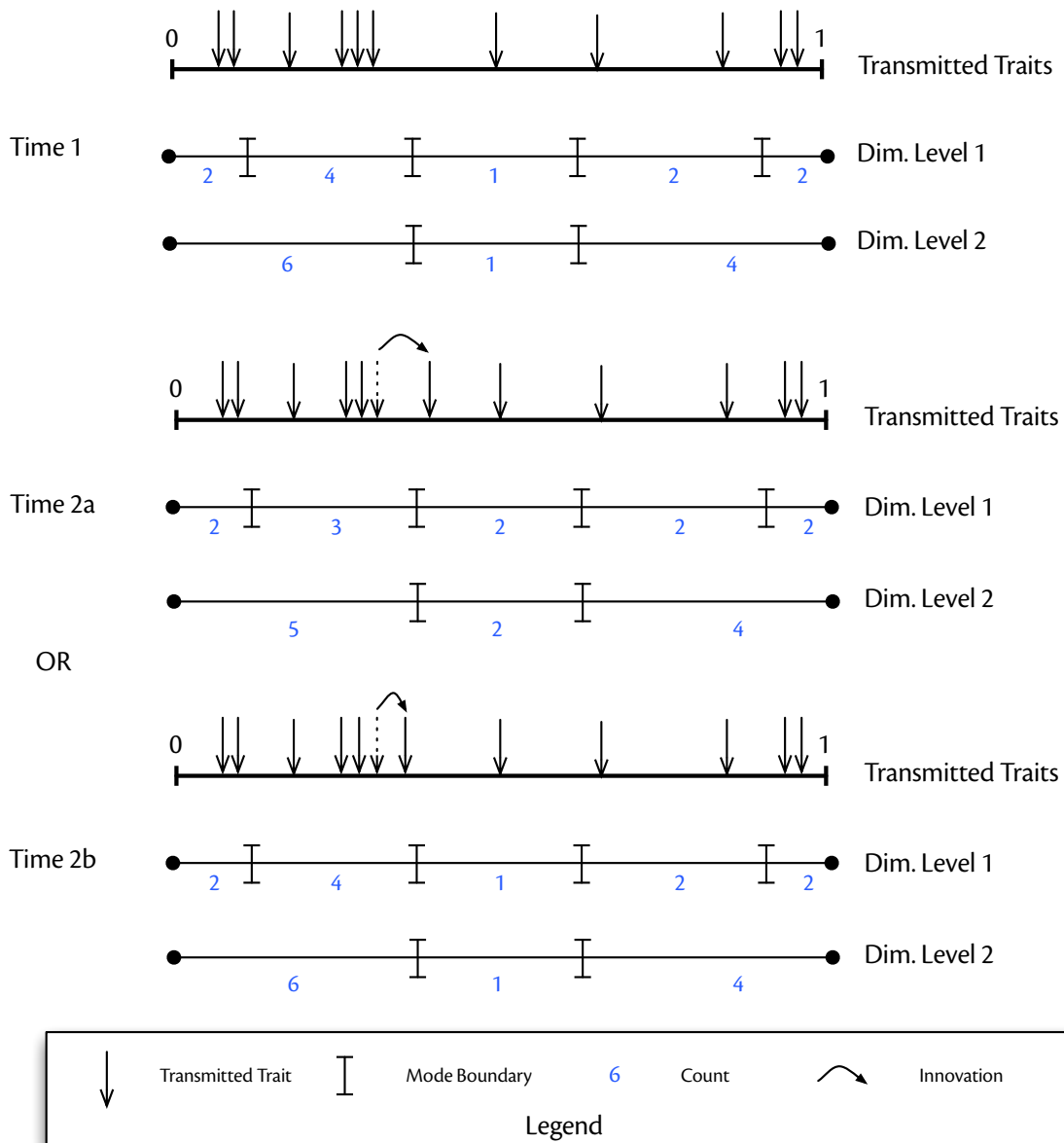


Figure A.8: Schematic view of observing transmission outcomes through analytic classes. Only one dimension of a paradigmatic class is shown for visual clarity, but normally more dimensions would be modeled. Transmitted traits are shown as points on the unit interval $[0, 1]$, with innovation creating a new unoccupied point on the interval, a random distance away from the trait being “mutated.” The outcome of this process is being monitored by two hierarchical levels of analytic class, labeled 1 and 2, with 1 nested inside the definitions of 2. The vertical lines with bars indicate the boundaries between modes and schematically indicate mode “definitions.” Mode definitions are chosen at arbitrary distances and are not equal in length. Counts are shown below each dimension of the traits which map to each mode. See Section 5.1.2 for discussion of the two innovation scenarios.

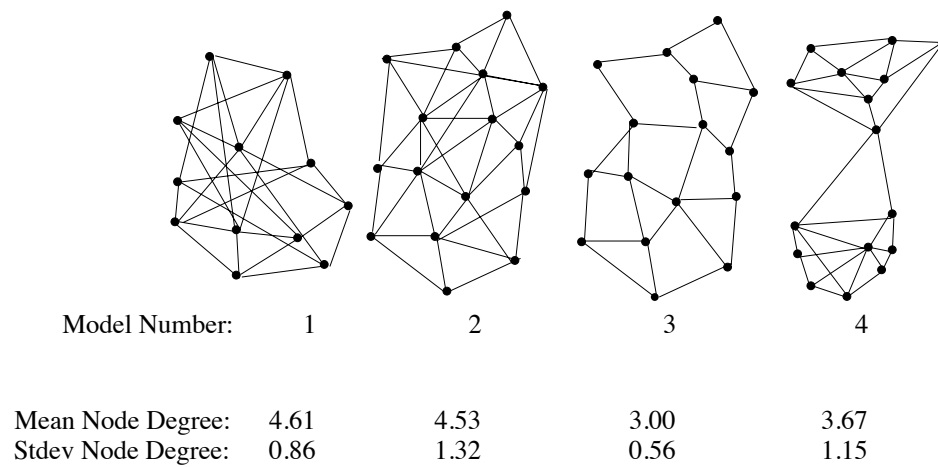


Figure A.9: Four simple models of transmission flow between demes in a metapopulation. The first model depicts long-distance connections and dispersal of information, and relatively high and even flow between demes, with no isolated or “hot” spots. The second model possesses no long-distance connections, but still relatively high and even flow between demes at short distances. The third model has “nearest-neighbor” connectivity, as is typically modelled in lattice or island models of population structure. The fourth model introduces variability in linkage between clusters of demes, creating what graph and network theorists call “community structure.” Reproduced with permission from [Lipo \(2001, Figure 4.10\)](#).

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 A.1: Comparison of expected K_n from Equation (6) 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.

Table A.2: 63 assemblages from Phillips et al. (1951) with more than 50 decorated sherds, tabulated with the number of decorated culture-historical types (richness). Total assemblage size and richness noted for comparison.

Site Name	Site Number	Total Sherds	Tot. Richness	Decorated Richness	Total Dec. Sherds
Parkin	11-N-1	6851	10	8	1303
Neeley's Ferry	11-N-4	7263	11	9	1269
Vernon Paul	11-N-9	2567	7	5	758
Williamson	11-N-13	4919	9	7	670
Leland	19-M-1	8394	12	10	647
Barton Ranch	11-O-10	4702	9	7	638
Silver City	20-O-5	2653	11	9	430
Castile	13-N-21	1999	11	9	394
Walls	13-P-1	3793	14	12	361
Powell Bayou	17-O-9	2230	17	15	351
Kinlock	19-N-1	5610	10	8	336
Arcola	20-M-1	2754	10	8	326
Cummins	11-O-4	713	8	6	261
Belle Meade	13-O-5	798	10	8	254
Rose Mound	12-N-3/A&B	1400	9	7	242
Kent Place	13-N-4/B,C,D,E	1026	9	7	241
Big Eddy	12-N-4	721	9	7	229
Hollywood	13-O-10	1486	13	11	226
Bush	17-M-11	4358	15	13	225
Stokes Bayou	16-M-6	2764	17	15	217
Fortune	11-N-15	685	5	4	200
Myer	16-N-10	1245	17	15	198
Jaketown	20-O-1	1157	9	7	197
L. Cormorant	13-P-8	930	12	10	192
Merigold	17-N-1	1903	18	16	183
Clay Hill	13-N-7	654	8	6	181
Wallace	17-K3	1507	9	8	176
Starkley	13-N-16	918	9	7	171
Mound Place	12-P-1	3178	12	10	168
Nickle	13-N-15/C,D,E	737	10	8	163
Carson Lake	10-P-1	3489	10	8	159
Montgomery	15-N-6	945	14	12	153
Owens	14-O-2	657	11	9	144
Deer Creek	20-M-2	1289	8	6	138
Cramor Place	12-O-5	1008	8	6	138
Turnbow	11-N-12	3052	9	7	129
Commerce	13-O-11	1097	10	8	128
Lipe	18-M-4	903	6	4	117
Parchman	15-N-5	998	13	11	114

Dupree	16-L-6	819	9	8	111
Woodlyn	13-P-11	847	10	8	107
Menard	17-K-1	1043	13	11	102
Grant Place	13-N-11	515	6	4	93
Irby	13-P-10	758	10	8	91
Salomon	15-O-1	1164	11	9	84
Alligator	16-N-2	1175	10	8	83
Marlow	18-N-1	486	8	6	81
Beck	13-O-7	257	8	6	79
Davis	13-N-5	456	7	5	79
Moore	14-N-1	296	7	5	78
Oliver	16-N-6	553	14	12	75
Oliver	16-N-6	552	11	9	74
Alma Brown	17-K-7/E/B	488	7	5	73
Old Town	15-N-3	506	8	6	66
Shelby Place	12-P-2	481	12	10	62
Dundee	14-O-8	621	7	5	61
Pouncey	12-O-2	439	7	5	60
West	14-O-10	258	8	6	57
Spendthrift	16-O-2	901	8	6	55
Vance	17-N-9	391	6	4	51
Notgrass	10-P-4	3182	13	11	50
Perry	14-O-4	321	8	6	50
Stoneville	19-M-3	229	8	6	50