Chapter 5
Analysing Cultural Frequency Data: Neutral Theory and Beyond

Anne Kandler and Enrico R. Crema

Introduction

The question of how humans use social information has been subject of a large amount of empirical and theoretical research in a variety of scientific disciplines (e.g. Coultas 2004; Mesoudi and O'Brien 2008; Kirby et al. 2008; Caldwell and Millen 2009; Henrich 2001; Bentley et al. 2004; Baum et al. 2004; Rendell et al. 2010; McElreath et al. 2008; Morgan et al. 2012). But in order to answer this question directly one would need fine-grained individual-level data detailing who learns from whom. However, outside of controlled experimental conditions, large longitudinal data sets of this kind are difficult to obtain (but see Henrich and Broesch 2011; Beheim et al. 2014), especially so in archaeological contexts. The archaeological record documents the frequencies of different cultural variants in sparse samples taken from the whole population for a single or multiple time intervals. As these frequency data often present the only direct empirical information about past cultural traditions and the forces affecting them (Shennan 2011), researchers have attempted to use the population-level patterns to infer processes of cultural transmission that may underlie them. But given the large number of transmission processes that have been identified in the literature (e.g. Laland 2004) and the sparseness of the archaeological record this represents a challenging task, especially as it is far from obvious whether the observed frequency data contain a strong signature about the underlying transmission processes.

Early approaches based on the pioneering work of Neiman (1995) addressed this challenge by using principles and methods derived from the neutral theory of molecular evolution (see, e.g., Shennan 2011; Eerkens and Lipo 2005; Lycett 2015, for comprehensive reviews, but see Chap. 7 in this volume for an alternative approach). Using mostly ceramic assemblages researchers tested whether the observed frequency distributions at particular points in time could be distinguished from the ones that may emerge under the hypothesis of unbiased transmission, i.e. in a situation where cultural variants are chosen at random for reproduction. In other words, this research attempted to establish whether observed data sets are consistent with neutral evolution. In the first part of this chapter we briefly

A. Kandler (✉)
Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
e-mail: anne_kandler@eva.mpg.de

E. R. Crema
Department of Archaeology, University of Cambridge, Cambridge, UK
e-mail: erc62@cam.ac.uk

summarise cultural neutral theory and evaluate what consistency or inconsistency between observed empirical patterns and neutral expectations, derived from commonly used statistical tests, can tell us.

In the second part we focus on the generative inference approach, recently developed in population genetics (e.g. Veeramah et al. 2011; Eriksson et al. 2012; Posth et al. 2016; Ray et al. 2009) which goes beyond the hypothesis testing framework. This approach simultaneously evaluates the consistency of a number of cultural transmission processes with the available data while also accounting for demographic and cultural properties of the system considered. In this way we can analyse the challenge of inferring underlying processes of cultural transmission from sparse population-level frequency data in more depth, in particular we can identify equifinal transmission processes (cf. Cochrane 2009; Premo 2010; Crema 2018; Kandler and Powell 2018), i.e. processes that are able to generate very similar population-level patterns, and therefore investigate how much information about underlying transmission processes can be extracted from sparse archaeological data.

Readers interested in running the simulation models presented in this chapter can install the dedicated R package (which can be found on https://github.com/ercrema/HERAChp.KandlerCrema and https://doi.org/10.5281/zenodo.1422010) and follow the vignette for reproducing the analysis presented here.

Neutral Theory

Introduced prominently by Kimura (1968, 1983) the neutral theory of molecular evolution claims, unlike the Darwinian theory of evolution by natural selection, that the overwhelming majority of evolutionary changes at the molecular level are not caused by natural selection acting on advantageous mutants but by random fixation of selectively equivalent or nearly equivalent mutants through the cumulative effects of sampling drift due to finite population size and mutation processes (Kimura 1991). Importantly, the relatively simple assumptions of neutral theory enabled the development of a large body of mathematical theories to treat molecular evolution and variation in quantitative terms. In part the success of neutral theory is rooted in the fact that these mathematical theories generate strong predictions that can be tested against data.

Subsequently other scientific disciplines started exploring the applicability of neutral evolution. Building on the foundational ideas, neutral theory in ecology, most prominently expressed in Hubbel's unified neutral theory of biodiversity and biogeography (Hubbell 2001), seeks to capture the influence of speciation, extinction, dispersal, and ecological drift on diversity and the relative species abundance under the assumption that all species are demographically alike on a per capita basis (Volkov et al. 2003). In particular, neutral theory implies that niche differences are not needed to explain biodiversity patterns. As in the population genetics case the assumption of neutrality allowed for the development of a tractable theory for testing hypotheses about community assembly rules (Volkov et al. 2003). In both fields the hypothesis of neutral evolution generated huge controversies which ultimately advanced our understanding about the evolutionary process.

Also in archaeology neutral theory has been used to analyse the composition of cultural assemblages. But already prior to that a lot of research had been dedicated to distinguishing "functional" from "stylistic" aspects of artefact traditions and to identifying the forces acting on them as they change through time (e.g. Binford 1963; Wiessner 1983). These discussions took a new direction when Dunnell (1978) proposed that "stylistic" aspects of artefacts could be defined as "those not under selection". Subsequently, Neiman (1995) used the framework of molecular neutral theory to introduce the idea that observed temporal changes in cultural assemblages can be explained by random processes involved in the neutral theory and the factors that affect these processes, such as effective population size or innovation rate. In particular, this model assumed that artefacts (or cultural variants in general) are chosen to be replicated according to their relative frequency, and new variants not previously seen in the populations are introduced by a process resembling random mutation (i.e. innovation). In finite

populations this copying process is affected by sampling error, and consequently at each iteration we expect some changes in the frequency of the variants. After multiple iterations frequencies will drift away from their original values, particularly when population size is small. If no variants are introduced via innovation, this iterative process will eventually lead a given variant to either go extinct or to fixate. However when innovations are introduced, this unbiased transmission process alone can replicate the rise and fall in the popularity of cultural variants over time as well as their disappearance. Neiman (1995) applied this theory to a data set recording the frequencies of selectively neutral stylistic elements in Woodland ceramics over time and showed that the patterns of within- and between-assemblage diversity in those stylistic elements could be explained by a hypothesis of unbiased and inter-group transmission. Following this pioneering work neutral theory has been applied to a number of archaeological case studies (e.g. Bentley et al. 2004; Shennan and Wilkinson 2001; Lipo 2001; Kohler et al. 2004; Eerkens and Lipo 2005; Schauer 2008; Steele et al. 2010).

In the following we provide some mathematical details on modelling neutral evolution (section "Mathematical Formulation: Wright–Fisher Model"), commonly used statistical tests for detecting departures from neutrality in cultural data (section "Detecting Departures from Neutrality") and their applications to archaeology (section "Application to Archaeology") as well as a discussion of how consistency or inconsistency between neutral expectations and data may be interpreted (section "Interpreting the Test Results").

Mathematical Formulation: Wright–Fisher Model

Neutral theory in cultural evolution has been mainly modelled using the Wright–Fisher infinitely many allele model (see, e.g., Ewens 2004, for a review of the mathematical properties). This model assumes that the composition of the population of cultural variants at time t is derived by sampling with replacement from the population of variants at time $t - 1$ resulting in non-overlapping generations. In more detail, the population of cultural variants at time $t - 1$ can be described by the abundances $[m_1, m_2, \ldots, m_{k_{t-1}}]$ of all k_{t-1} variants present at this time. It holds $\sum_{i=1}^{k_{t-1}} m_i = N$ where N denotes the temporally constant population size. In order to generate the population of cultural variants at time step t, N copying events are carried out. In each of these copying events a variant i is randomly chosen from the population at time $t - 1$, i.e. with probability

$$\pi_i = \frac{m_i}{N}(1 - \mu), \tag{5.1}$$

and a new instance of variant i is produced. Consequently a new population of cultural variants with the abundances $[n_1, n_2, \ldots, n_{k_t}]$ is generated. The term m_i/N (we later refer to it as p_i) describes the relative frequency of variant i in the population, i.e. Eq. (5.1) means that a variant is chosen to be reproduced proportional to its relative frequency and therefore the transmission process is denoted as unbiased. The variable μ stands for the innovation rate, meaning that with probability μ a new cultural variant, not currently or previously seen in the population, is introduced in each copying event. Repeating this process for many time steps will generate the rise and fall in the popularity of cultural variants over time, similar to diachronic patterns in type frequency observed by cultural historians (see Fig. 5.1).

Importantly, if the innovation rate μ is larger than 0, then every cultural variant will eventually go extinct. It may take a long time but it will happen with probability 1. This implies that variant frequencies will never stabilise over time but after sufficiently many time steps the cultural system will reach a stationary state where some aspects of system do not change anymore (see Box 5.1 for mathematical details). For instance, at steady state the expected level of cultural diversity at the
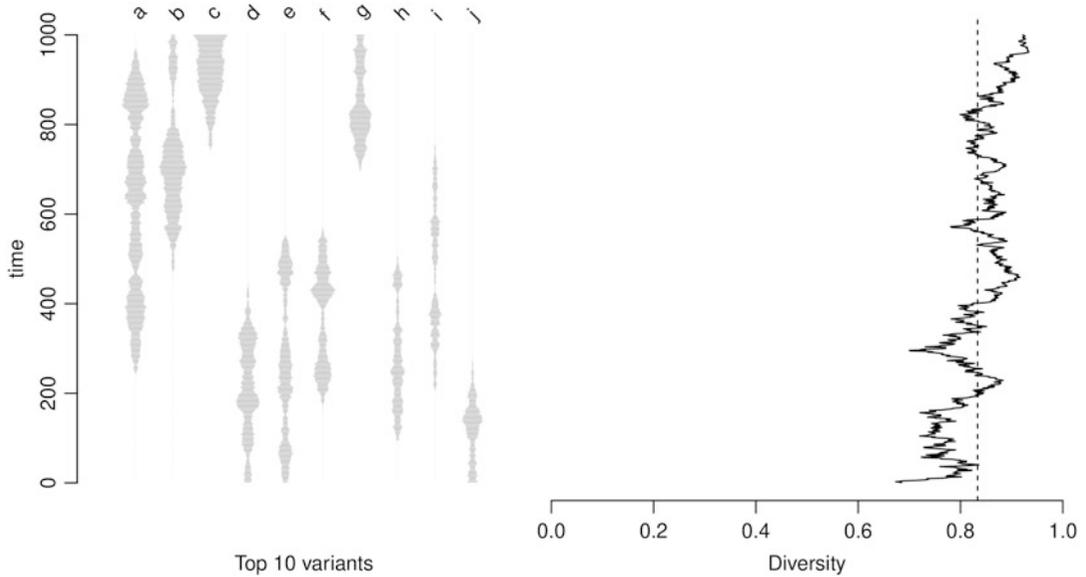
Fig. 5.1 Left figure: battleship curve of the ten most common variants across 1000 time steps obtained from a simulation run of unbiased transmission with $N = 500$ and $\mu = 0.005$. Right figure: corresponding time-series of the level of cultural diversity calculated by $1 - \sum_{i=1}^{k_t} p_i^2(t)$ (solid line) where k stands for the number of variants present and neutral prediction according to Eq. (5.4) (dashed line). It is obvious that the time series fluctuates around the neutral prediction

population level or the expected number of different cultural variants in a sample of size n (with $n << N$) can be derived under the assumption of neutrality (see Eqs. (5.4) and (5.6) in Box 5.1 and Fig. 5.1).

Box 5.1: Mathematical Details of the Wright–Fisher Model

The properties of the infinite allele Wright–Fisher model are well-understood and we summarise here only some of the characteristics (see, e.g., Ewens 2004, for more details). The probability that the population of cultural variants with abundances $[m_1, m_2, \ldots]$ at time $t-1$ is transformed into a population with abundances $[n_0, n_1, n_2, \ldots]$ at time t (where n_0 describes the abundance of all innovations in this time step) is given by

$$P(X_0(t) = n_0, X_1(t) = n_1, \ldots | X_1(t-1) = m_1, \ldots) = \frac{N!}{\prod_i m_i!} \prod_i \pi_i^{n_i} \qquad (5.2)$$

with $\pi_0 = \mu$, $\pi_i = (m_i/N)(1-\mu)$, and $\sum_i m_i = \sum_i n_i = N$. The state space of the Markov process defined by these transition probabilities can be extremely large making the derivation of population-level properties of this stochastic process almost intractable. But as the neutrality assumption implies that all variants are considered identical, the time evolution of a single variant can be described by a two-variant formulation

$$P(X_i(t) = n_i | X_i(t-1) = m_i) = \binom{N}{n_i} \pi_i^{n_i} (1 - \pi_i)^{N-n_i}. \qquad (5.3)$$

(continued)

Box 5.1 (continued)

Importantly, for $\mu > 0$ Eq. (5.2) suggests that the extinction of any variant is inevitable over time and consequently there does not exist a non-trivial stationary distribution for the variant frequencies. Nevertheless, it has been shown that some stationary properties of a cultural system evolving through neutral evolution can be determined. At steady state, it holds that the level of cultural homogeneity (defined as the probability that two instances randomly drawn from the population are of the same variant) at the population level can be approximated by

$$F \approx \frac{1}{1+\theta} \quad \text{with } \theta = 2N\mu. \tag{5.4}$$

The corresponding level of cultural diversity is given by $1 - F$. Further, for random samples of known size n (with $n << N$) the probability that k different variants are observed is given by

$$P(K = k) = \frac{|S_n^k|\theta^k}{S_n(\theta)} \tag{5.5}$$

where $S_n(\theta) = \prod_{i=0}^{n-1}(\theta + i)$ and $|S_n^k|$ is the absolute value of a Stirling number of the first kind. Derived from the expression above, the expected value of the number of observed variants in a sample of size n has the form

$$\mathbf{E}\{K\} = \sum_{i=0}^{n-1} \frac{\theta}{\theta + i}. \tag{5.6}$$

If precisely k variants are observed in the sample, the probability of an assemblage $[n_1, \ldots, n_k]$ of cultural variants is given by

$$P([n_1, \ldots, n_k]|k, n) = \frac{n!}{|S_n^k|k!n_1 n_2 \cdot \ldots \cdot n_k} \tag{5.7}$$

which is known as Ewens sampling distribution (Ewens 1972). Further, the probability that a variants with only one instance are observed in a sample of size n and k variants is

$$P(A_1 = a|k, n) = \sum_{j=a}^{k-1}(-1)^{j-a}\frac{|S_{k-1}^n|}{a!(j-a)!|S_k^n|} \tag{5.8}$$

which is approximately Poisson distributed with mean $|S_{k-1}^n|/|S_k^n|$ (Ewens 2004).

Archaeological applications of the Wright–Fisher model have mainly focused on two issues: inference of temporal changes in the values of the population size N (or effective population size) or the innovation rate μ from the observed frequency distribution of variants when there is no departure from neutrality assumed; and inference of non-neutrality (i.e., the existence of processes other than unbiased transmission) at a given location. In the following we focus on the latter application and discuss three approaches to detect departures from neutrality. Before delving into this statistical aspect we briefly mention an alternative framework for modelling neutral evolution that allows for overlapping generations but is only rarely used in archaeological studies.

Moran Model

The classical Wright–Fisher model described above assumes non-overlapping generations, i.e. after one time step the whole population of cultural variants is assumed to "die" and to be replaced by an offspring generation. This assumption is often at odds with reality. The model by Moran (1958) allows for overlapping generation, i.e. there is variation in the individual life spans of the instances of a variant. In more detail, the Moran model assumes that in each time step only one instance of the cultural variants is chosen at random to reproduce (whereby the reproduction/copying process is only faithful with probability $1 - \mu$ and, as above, with probability μ an entirely new variant is introduced). After reproduction one instance (excluding the new offspring instance) is chosen to "die". This model is an example of birth and death models which are studied extensively in the stochastic process literature. As with the Wright–Fisher model for $\mu > 0$ there is no concept of stationarity of the frequency of any cultural variant. However, the stationary distribution of variant configurations (and related results similar to the ones shown in Box 1) can be found (see, e.g., Ewens 2004).

Detecting Departures from Neutrality

An advantage of developing a mathematical framework of neutral theory is that it allows for the derivation of expectations about the values of certain statistics such as the expected number of distinct variants in a sample of size n which then can be compared to observed data. If those theoretical expectation and empirical patterns coincide, we can conclude that the hypothesis of neutral evolution is consistent with the data. In the following we briefly review three different approaches commonly used in cultural evolution literature for detecting departures from neutrality.

Ewens Sampling Distribution

One of the earliest approaches for testing the neutral hypothesis has been based on the cultural homogeneity index, defined analogously to its genetic counterpart as the probability that two randomly drawn cultural variants are of the same type. The Ewens-Watterson test of homozygosity (Ewens 1972; Watterson 1977, 1978) evaluates the empirical homogeneity statistic

$$f = \sum_{i=1}^{k} p_i^2 \tag{5.9}$$

where p_i denotes the relative frequency of variant i. To determine how small or large f has to be so that the hypothesis of neutrality is rejected, its expectation under neutrality or more precisely its neutral distribution has to be calculated. To do so, Watterson (1978) suggested an approximate procedure which randomly draws samples from Ewens sampling distribution (5.7) given the observed sample size n and the number k of different variants present in this sample and determines the corresponding homogeneity statistic. In this way reliable statistical estimates can be made of various significance level points. Thus the homozygosity test was designed to test against processes that lead to significantly lower or higher levels of diversity than predicted under neutrality.

In contrast, Slatkin's test (Slatkin 1994, 1996) does not employ the concept of homozygosity, and relies only upon the "shape" of Ewens sampling distribution. It evaluates the empirical sample against all possible configurations $[n_1, \ldots, n_k]$ for given n and k drawn from distribution (5.7) and therefore makes fewer assumptions about the characteristics of the distribution that may indicate the presence of selective forces.

Both tests have been applied by archaeologists, mainly to analyse the composition of ceramic assemblages. The homozygosity test was explicitly or implicitly used by, e.g., Lipo (2001), Shennan and Wilkinson (2001) and Kohler et al. (2004). The Slatkin's exact test was used by, e.g., Steele et al. (2010), Premo and Scholnick (2011) and Bortolini (2014).

Progeny Distribution

Taking a different view, Bentley and colleagues (e.g. Bentley et al. 2004; Hahn and Bentley 2003; Herzog et al. 2004) used the Wright–Fisher model to explore the properties of the neutral progeny distribution. The progeny distribution records the frequencies of cultural variants which produce k new variants over a fixed period of time (We note that this distribution naturally accounts for the problem of time-averaging.). Through simulation techniques (e.g. Bentley et al. 2004; Hahn and Bentley 2003; Premo 2014) they concluded that the neutral progeny distribution takes the form of a power law. The exponent, $\alpha(N, \mu)$, of this power law has been fitted as a function that depends on innovation rate μ and total population size N. Consequently, they hypothesised that testing whether an empirical progeny distribution describe a power law with an exponent similar to the 'neutral' exponent α for given N and μ allows for conclusion about the consistency between data and neutral evolution. This method has been applied to data sets describing the choice of baby names in the US, US patents and their citations or Neolithic pottery motifs and provided support for the neutral hypothesis in all cases (Bentley et al. 2004; Hahn and Bentley 2003).

However, we note that results of this analysis have to be interpreted cautiously as their accuracy depends on the completeness of the data set considered. To illustrate this O'Dwyer and Kandler (2017) used an overlapping generations model commonly applied in ecology and derived an analytical representation of the neutral progeny distribution. In agreement with the earlier work, it was shown that neutral theory generates a power-law progeny distribution but with a constant exponent of 3/2 (i.e. the power-law exponent does not depend on innovation rate or population size). The power law is followed by an exponential cut-off, whereby the onset of this cut-off depends on the innovation rate: the larger the rate, the earlier is the onset. The analytical representation of the progeny distribution allowed for maximum-likelihood estimations of the model parameter and therefore provided a direct way of parameterising neutral models using cultural data, and of subsequently evaluating the consistency between observed data and the neutral hypothesis (The code of the framework can be downloaded from https://github.com/odwyer-lab/neutral_progeny_distribution). Importantly, this study established that analyses based on only the most popular variants, as is often the case in studies of cultural evolution, can provide misleading evidence for neutral evolution. If the data set is incomplete, i.e. if it does not include rare variants, then consistency of neutral theory with the empirical patterns can be inferred even in situations where selection processes are acting in the population. While the analysis of a complete data set recording boys and girls first names in Southern Australia revealed a mismatch between neutral theory and empirical data, the analysis of the same data set but with all names possessing abundances of less than five in the considered time interval removed showed consistency between neutral theory and empirical data (O'Dwyer and Kandler 2017). This result points to the crucial importance of rare variants for evaluating processes of cultural evolution based on aggregated population-level data in the form of progeny distributions.

Turn-Over Rates

Focusing on the temporal dynamic of cultural change, Bentley et al. (2007) analysed the turn-over behaviour in top lists by determining the number of new variants to enter the list of the y most popular/frequent variants in the population relative to the previously sampled interval (see ElBahrawy

et al. 2017, for an alternative definition). Based on simulations of the Wright–Fisher model they estimated that under neutral evolution the average number of turn-overs for a top-list of length y (denoted as z_y) is given by

$$z_y \approx y\sqrt{\mu}. \tag{5.10}$$

The turn-over rate seems to be approximately proportional to the list length and largely independent of the underlying population size. These theoretical predictions have been compared to estimations of turnover-rates for the Billboard Top 200 Pop Chart, various top lists for US baby names, dog breeds (Bentley et al. 2007) and cryptocurrencies (ElBahrawy et al. 2017) and a good coincidence has been obtained.

A subsequent study by Evans and Giometto (2011) revealed that this relationship can be more accurately described by

$$z_y = A \cdot y^x \tag{5.11}$$

where the coefficient A depends on N. If it holds $N\mu > 0.15y$, then the exponent x assumes the value of 0.86 under neutrality (Acerbi and Bentley 2014) and thus empirical estimates of x can be used to identify instances where the observed turn-over profile deviates from the patterns expected under neutrality. Acerbi and Bentley (2014) explored this idea and examined the shape of the turn-over profile and estimates of x under three alternative biased transmission processes (neutral evolution, negative and positive frequency-dependent selection). They concluded that in some situations the estimate of x is indicative of the underlying process of cultural transmission.

In general, however, the turn-over rate approach has limited applicability for archaeological data, and there are currently no known applications. The coefficient x can be estimated by regression based techniques but these estimates are sensitive to the number of data points available to calculate the turn-over rate. To illustrate this point Fig. 5.2 shows the interquartile range, i.e. the range between the 25th and 75th percentile, of the estimated values of x from simulation data generated under neutral model with fixed values for μ and N, but different number of time steps from which turn-over rates were calculated. The graphs show that with increasing number of time steps the variance in the estimates of x becomes smaller. Data sets with less than 40 times-steps—as typically encountered in archaeology—are likely to yield estimates of x that strongly diverge from the theoretical expectations (displayed as dashed line in Fig. 5.2) even when the underlying process is neutral evolution.

Alternative Approaches

Stronger tests of departure from neutrality than testing the goodness of fit between empirical and theoretical frequency distributions have been developed mainly in population genetics and ecology. These tests often involve the analysis of the temporal dynamic of change. For example, it has been shown that neutral theories in ecology have had less success in predicting the dynamics of biodiversity, from decadal-scale species abundance fluctuations to geological ages of species (Leigh 2007; Wang et al. 2013; Chisholm and O'Dwyer 2014; O'Dwyer et al. 2015).

Similarly, recent work in cultural evolution has pointed to the importance of analysing temporal patterns of change as opposed to static measures of cultural diversity (e.g. McElreath et al. 2005; Hoppitt et al. 2010; Kandler et al. 2017; Kandler and Shennan 2013; Sindi and Dale 2016). For example, Sindi and Dale (2016) analysed the patterns of frequency change, in particular, the kurtosis of the distribution of changes over time, of stable words in the Google Ngram database. Interestingly, this approach identified words under selection: kurtosis values close to zero signalled neutrality while deviations from zero were indicative of selection. Approaching this problem from a different
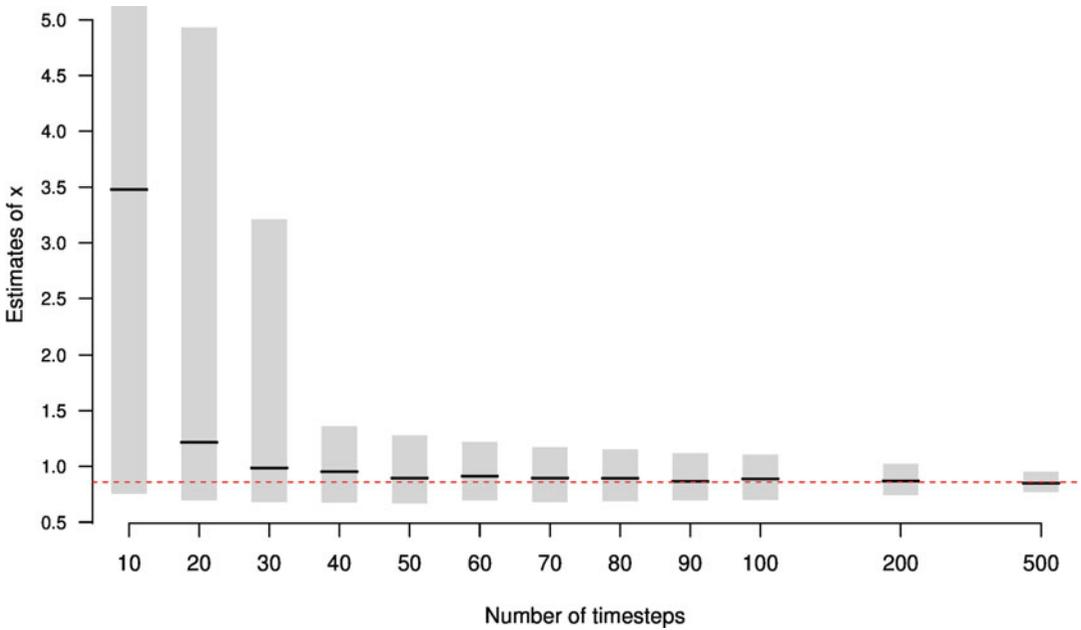
Fig. 5.2 Interquartile ranges of estimates of x based on different numbers of time steps. The exponent x and the coefficient A in Eq. (5.11) are estimated by fitting a linear model to $ln(z_y)$ using $ln(y)$ as predictor variable. The coefficient of the predictor represents an estimate of x and the intercept the estimate of A. The response variable z_y is however the *average* turn-over rate, which will fluctuate considerable for smaller number of time steps. The grey areas show the interquartile range of the estimated values of x obtained from 1000 simulations of unbiased cultural transmission with $N = 500$ and $\mu = 0.01$ but different number of recorded time steps. The red dashed line illustrates the theoretical expectation of x equivalent to 0.86

angle, Steele et al. (2010) showed, using the example of Hittite ceramic bowl types, that the neutral hypothesis was not rejected on the basis of the frequency distribution. However, examination of the characteristics of the bowl types revealed a correlation between those characteristics and their abundance ranking which indicated clear departure from the neutral hypothesis (which assumes functional equivalence and therefore predicts the absence of such a correlation). Brantingham and Perreault (2010) analysed the same data set and extended the use of the Price equation to the analysis of selective and stochastic forces operating on multiple artefact types within an assemblage. Based on this approach they concluded that selection is a dominant process driving the frequency evolution of the different bowl types within the assemblage and that stochastic forces played little or no role.

Application to Archaeology

Archaeological applications of the methods described above have been primarily focused on ceramic assemblages (but see Bortolini 2014; Scholnick 2010, for exceptions). Most of these case studies rest on the comparison between expected and observed homogeneity index derived from Ewens sampling distribution (e.g. Neiman 1995; Shennan and Wilkinson 2001; Lipo 2001; Kohler et al. 2004; Schauer 2008; Steele et al. 2010; Premo and Scholnick 2011; Bortolini 2014; Scholnick 2010).

Methods based on the progeny distribution have been applied only to the Merzbach assemblage mentioned above, yielding contrasting results (Bentley et al. 2004). While Shennan and Wilkinson's original study (Shennan and Wilkinson 2001) suggested the rejection of the neutral model on the

grounds of a level of cultural diversity lower than that expected under neutrality, Bentley et al. (2004) argued that this might be the case for earlier phases but not for the later ones where the observed pattern have a good fit with the expected neutral progeny distribution (Bentley and Shennan 2003).

As mentioned earlier, we are unaware of any application of the turn-over rates based method, most likely because its strict requirement of frequency data recorded across a relatively larger number of time points for accurate estimates of turn-over rates (cf. Fig. 5.2).

The limited number of time points and phases is only one of the many issues challenging the application of these methods to archaeological assemblages. Premo (2014), for example, explored the problem of time-averaging and showed how methods based on diversity indices have an inflation of type I error with increasing duration (hence magnitude of time-averaging) of archaeological phases. This is simply due to the fact that with increasing duration more variants with shorter life spans are likely to be included, effectively increasing the observed number of variants and cultural diversity of the assemblage. Premo (2014) noted that progeny distributions are more robust to the effect of time-averaging.

Another key issue in the applications of these methods is the exact definition of what constitutes a cultural variant. All the methods summarised above require count data, but counts of what? In some domains such as baby names or dog breeds this is arguably a straightforward exercise (though one may argue whether *Cliff* and *Clifford* should be regarded as instances of the same variant or two distinct ones), but in other contexts this is less so. Some decorative traits in pottery design might be affected by substantially high levels of inter-observer variability in defining what constitutes the unit of transmission (or replicator), potentially biasing the inferential enterprise. This issue is also linked to the extent by which the units of transmission can be assumed to be discrete entities rather than continuous ones. While the extent of the problem is, again, likely to vary case by case, theoretical models have shown that in presence of strong cognitive attractors, discrete replicator approximation can be effective even if the unit of transmission is continuous (Henrich and Boyd 2002). It is worth pointing out that the question of what constitute a cultural trait is also epistemological in nature, and can be described as a classification issue (see also Lipo et al. 1997; Pocklington and Best 1997; O'Brien et al. 2010).

Lastly, in many contexts changes in the frequency of specific variants may be strongly affected by the mode of production and extent of unevenness in productivity. For example, an assemblage of n ceramic bowl produced by only two potters engaged in serial production may exhibit a different pattern of cultural variation than an assemblage of the same size but produced by, e.g., ten potters. While these issues of cultural growth *vs* cultural reproduction have been discussed from a philosophical standpoint (see, e.g., Ramsey and De Block 2017), their potential inferential challenges, particularly in production economies, have yet to be explored adequately.

Interpreting the Test Results

At a first glance, the assumptions of neutral theory are often at odds with the vast stores of knowledge archaeologist and anthropologists have accumulated for social systems. Humans are generally not thought of as making decisions at random. Neutrality would imply that individuals do not possess any preferences for existing cultural variants nor does the adoption of a particular cultural variant provide an evolutionary advantage over the adoption of a different variant. Additionally neutral evolution assumes that each cultural variant evolves independently. While these inherent assumptions are likely to be violated (for detailed discussions see, e.g., Neiman 1995; Shennan and Wilkinson 2001; Steele et al. 2010), it has been shown that population-level patterns of various observed episodes of cultural change resemble the ones expected under neutrality (see, e.g., Bentley et al. (2004); Hahn and Bentley (2003); Neiman (1995); Bentley et al. (2007) and section "Application to Archaeology"). In other
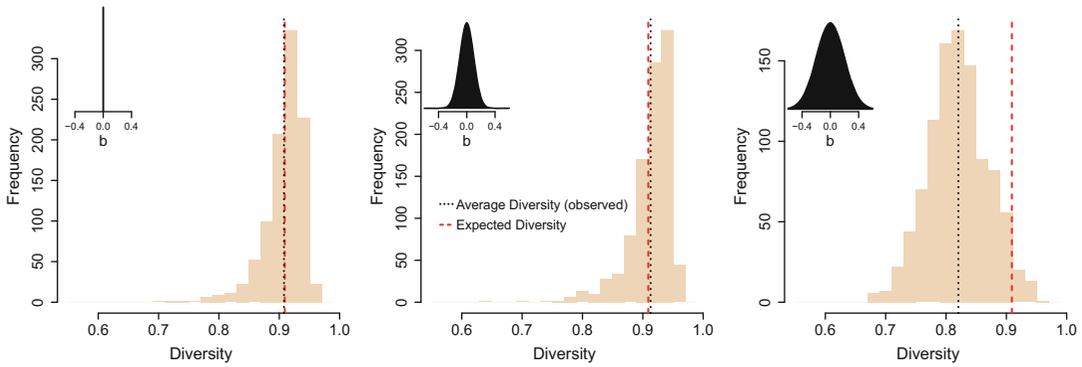
Fig. 5.3 Distribution of the level of cultural diversity for increasing levels of heterogeneity in individual preferences after 1000 time-steps with $N = 500$ and $\mu = 0.01$. We assumed that individuals can have different propensities for frequency-dependent transmission, i.e. individuals can possess different b values (see Eq. (5.12)), but kept the mean propensity of the population at zero which would imply unbiased transmission. Left figure: all individuals use unbiased transmission (i.e. $b = 0$); centre figure: individual propensities at each time step are drawn from a normal distribution with mean 0 and standard deviation 0.1 (see inset); right figure: same situation as in the central figure but with standard deviation increased to 0.2 (see inset). The black dashed lines represent the mean level of cultural diversity across all simulations and the red dashed lines the neutral predictions $1 - F$ (see Eq. (5.4)). If the deviations of the propensities from zero are not too large (central figure), then the resulting population-level patterns are consistent with neutral theory (cf. left figure and dashed red line). If, however, the deviations get larger (right figure), then the population-level patterns deviate from neutral expectations and consequently there is no guarantee that a population expressing heterogeneous preferences will exhibit a population-level signal of neutrality

words, we are confronted with a situation where we *know* that individuals possess heterogeneous preferences for choosing one cultural variant over another, however, the resulting population-level patterns may still be consistent with neutrality.

So what does consistency between population-level predictions of the neutral theory and empirical data tell us? It has been suggested that if each individual act of choosing one cultural variant rather than another has a different motivation, the emerging population-level patterns will be that there are no directional selective forces affecting what is copied, and therefore population-level patterns of cultural change should be consistent with neutral theory (e.g. Shennan 2011). While this is plausible at least for narrow distributions of individual preferences (but see Fig. 5.3 which shows that there is no guarantee for a population-level signal of neutrality if individual preferences are very heterogeneous) we need to be careful with the "reverse" argument that the presence of collective selective forces or systematic biases in the transmission process will produce patterns that deviate from neutral expectations, which in turn would make neutral theory a suitable cultural null hypothesis (see, e.g., Richerson and Boyd 2008). Although this interpretation is very appealing we discuss two potential problems in the following (for discussions in the ecological context see, e.g., Gotelli and McGill 2006; Rosindell et al. 2012).

Link Between Population-Level Patterns and Individual-Level Processes

On the one hand, the cultural evolution literature has identified a large number of selective cultural transmission processes (e.g. Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985; Laland 2004). Henrich and McElreath (2003) categorised these selective biases into two broad groups: content biases and context biases. While content biases refer to differential copying or replication of a cultural trait due to outward features it displays (e.g. a cultural variant provides a functional benefit compared to other alternatives), context biases do not refer directly to physical features of cultural variants but

the social or material context in which they are used (e.g. a cultural variant is copied due to its high frequency in the population; Lycett 2015). On the other hand, archaeological data sets usually record frequencies of different variants of cultural artefacts in a sample taken from a population of often unknown size at a single or several time points and therefore possess a sparse nature. Consequently, we face a scenario where there are many transmission hypotheses that could potentially explain the observed (sparse) data and we should not a priori expect a unique relationship between underlying processes of cultural transmission and population-level frequency patterns. In other words, we need to account for the problem of equifinality, i.e. situations where various individual-level processes can result in very similar population-level characteristics (e.g. von Bertalanffy 1969; Premo 2010). The problem is even worsened in populations of small sizes where the effects of random drift are magnified potentially obscuring any signatures of selective processes in the data. More generally random drift and innovation are not exclusive to neutral theory and hence expected differences in population-level patterns are rooted in the specifics of the underlying cultural transmission process.

Kandler et al. (2017) started examining this problem theoretically by developing a simulation framework aimed at exploring the level of distinguishability of processes of cultural transmission from population-level statistics. To do so they generated the probability distributions for a number of statistics such as the level of cultural diversity conditioned on different transmission processes. Those distributions describe the possible range of values of the statistics that can be assumed under the same environmental conditions. Therefore the area of overlap between these distributions indicates to what degree the two corresponding transmission processes can be distinguished based on the statistic (without knowledge of an empirical estimate). At one extreme, no overlap suggests that the processes can be reliably distinguished (as they result in distinct values of the statistic); at the other, complete overlap suggests that they cannot be distinguished as they result in almost similar values of the statistic (see also Crema et al. 2014, and their use of confusion matrices to assess the extent of equifinality). Further, they calculated the probability that an empirical estimate of a particular statistics could have been generated by a transmission process x as opposed to another process y. The analysis revealed that different processes of cultural transmission (vertical, oblique, horizontal and unbiased transmission as well as their frequency-dependent versions) will result in very similar population-level patterns, especially if those patterns describe the cultural composition of a population at a single point in time. In other words, neutral evolution is likely not distinguishable from alternative transmission processes based on, e.g., the observed level of cultural diversity as measured by Eq. (5.9). However, the temporal dynamic of cultural change, as, e.g., measured by the time a variant stays the most common variant, can retain a stronger signature of the underlying transmission processes than a "snapshot" of the relative frequencies of the variants at a given point in time. These results suggest that even when outcomes are similar in terms of cultural composition, they can differ substantially in temporal dynamics: similar distributions of cultural variants at a specific point in time can be reached through substantially different processes. Therefore attempts to infer processes of cultural transmission from population-level data should be based on measures of the temporal dynamic of cultural change.

But it is often not possible to increase the temporal resolution of archaeological data sets to the level that would be needed to reliably distinguish between a number of cultural transmission processes. Therefore the consistency between empirical data and neutral evolution (i.e. unbiased transmission) needs to be interpreted carefully as other cultural transmission hypotheses may be equally consistent.

Equilibrium Assumption

Most of the approaches mentioned in section "Detecting Departures from Neutrality" for detecting departures from neutral theory rest on the equilibrium assumptions, i.e. the predictions only hold for cultural systems at steady state. This implies that the observed frequency data was generated from

a system where, among others, population size, innovation rate, and transmission process (unbiased transmission in the case of neutral evolution) were constant for a sufficiently long amount of time so that observed summary statistics are approximately constant over time and the initial conditions of the system play a minimal or no role. If this is not the case, e.g. because the population has recently undergone or is undergoing an expansion event or bottleneck, then those predictions are likely to be violated even if the population evolves neutrally. For example, in the situation of a bottleneck where the population size has decreased from N_1 to N_2 the level of cultural diversity "moves" from one steady state conditioned on N_1 to another steady state conditioned on N_2. However, this process takes time and any level of diversity observed in this transition period will not match the neutral prediction for neither N_1 nor N_2 (see also Rorabaugh 2014, for an extensive discussion of the effect of population bottleneck with regard to continuous neutral traits).

Figure 5.4 illustrates another potential source for non-equilibrium dynamics. Based on Wright–Fisher simulations with temporally constant population size N and innovation rate μ we show in panel (a) the level of cultural diversity as determined by $1 - f_t = 1 - \sum_{i=1}^{k_t} p_i^2(t)$ at each time step. The grey lines show the time course of the diversity level for a single simulation and the red-shaded area indicates the 95% prediction interval of the simulation output. The solid red line represents the corresponding average level of diversity at each time step and the dashed black line the neutral expectation $1 - F$ (see Eq. (5.4)). It is obvious that both, theoretical and simulation results match very closely.

In panel (b) we show the same results for a situation where the underlying process of cultural transmission changes for a brief time period (indicated by the blue shaded area) from unbiased transmission to negative frequency-dependent transmission. Negative frequency-dependent transmission is defined as the disproportional support for rare variants and is known to increase cultural diversity,
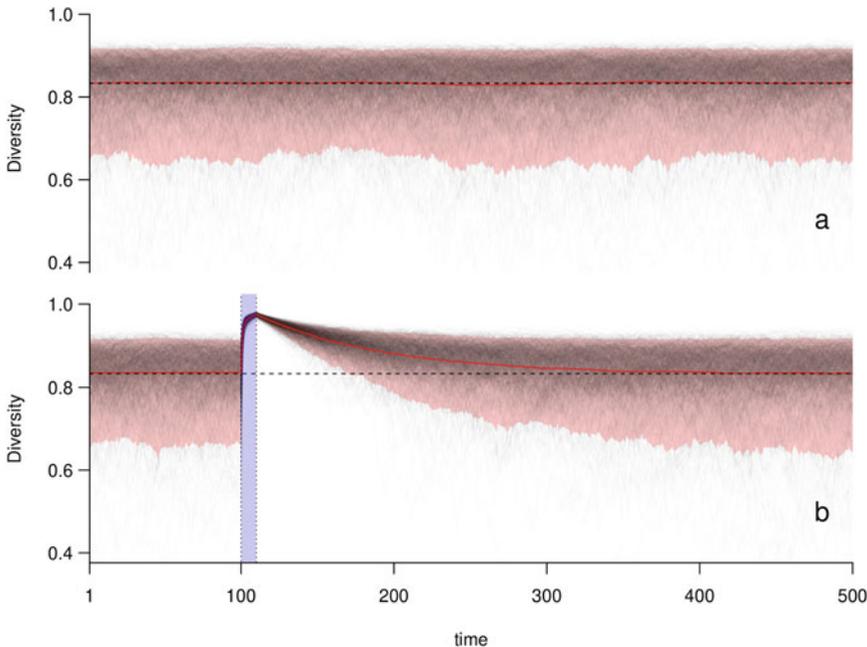

Fig. 5.4 Time series of the level of diversity under (**a**) equilibrium and (**b**) non-equilibrium regimes. In both figures the thin grey lines show the time series for individual simulation runs, the solid red lines the average diversity values, the red shaded area the 95% prediction interval out of 1000 simulations, and the black dashed lines the neutral expectation $1 - F$ (see Eq. (5.4)). The blue region in (**b**) highlights the interval where the transmission process temporarily changed to negative frequency-dependent transmission as described by Eq. (5.12) with $b = 0.5$

in particular to result in more even frequency distributions (Boyd and Richerson 1985). In modelling terms, this means that the probability that a variant i is chosen to be copied is changed from unbiased transmission expressed by Eq. (5.1)

$$\pi_i = \frac{m_i}{N}(1 - \mu)$$

to negative frequency-dependent transmission expressed by

$$\pi_i = \frac{\left(\frac{m_i}{N}\right)^{1-b}}{\sum\limits_{j=1}^{k}\left(\frac{m_j}{N}\right)^{1-b}}(1 - \mu) \quad \text{with } b > 0 \tag{5.12}$$

for a short time period. The term m_i/N represents the relative frequency of variant i in the population, μ the innovation rate, k the number of different variants present in the population and the coefficient b controls the strength of the frequency-dependent transmission ($b = 0$ results in unbiased transmission, $b < 0$ in positive frequency-dependent transmission and $b > 0$ in negative frequency-dependent transmission). We set $b = 0.5$ in our example. It is obvious that the level of diversity quickly reacts to the change in transmission process and increases. But after transmission returns to its unbiased way it takes time until the level of diversity reaches its equilibrium level again (see red solid line in panel (b) which needs several hundred time steps to reach the neutral expectation shown by the black dashed line). As in the situation of changing population size, if the level of diversity is measured in this transitional period, then the hypothesis of neutral evolution will be rejected as the empirical estimate does not coincide with the neutral expectation. Consequently, non-equilibrium dynamics may be entirely responsible for the discrepancy between empirical observations and theoretical expectations. In this context, Crema et al. (2016) argued that the steady state assumption should be a hypothesis to be tested, rather than simply held a priori.

Summarising, we need to be careful with interpreting the consistency (or inconsistency) of archaeological data with theoretical neutral expectations as evidence for the absence (or presence) of selective forces. While it is likely that selective processes of cultural transmission generate population-level patterns similar to the ones expected under neutrality, especially in the light of sparse data, deviations between data and neutral theory can be entirely caused by non-equilibrium dynamics resulting from, e.g., temporal changes in population size. Additionally, the observed composition of the archaeological record may be the product of the accumulation of variants from various time points what also can produce a discrepancy with neutral theory (Premo 2014; Madsen 2012; Perreault 2018). Given these problems, Richerson and Boyd (2008) suggested to fit all plausible cultural transmission hypotheses to the data and see which, if any particular one, fits best. In the next section we describe a modelling framework which follows this suggestion and has the potential to address some of issues discussed in this section.

Inferring Alternative Processes of Cultural Transmission

If our aim is to uncover the specifics of the cultural transmission process from population-level frequency data, then we should analyse the temporal dynamic of cultural change. Further, it would be desirable to develop a modelling framework that simultaneously evaluates the consistency of a number of transmission hypotheses with the available data while also accounting for demographic and cultural properties of the system considered.

The general problem we are facing—the inference of underlying evolutionary processes from population-level data—is of course not unique to archaeology. In fact, other scientific fields have successfully overcome similar problems, in particular population genetics, which aims to understand the evolutionary mechanisms that produced the allele frequency distributions observed both now and in the past. Recent work has focused on developing efficient generative inference frameworks, which allowed for the statistical testing of increasingly realistic and complex evolutionary hypotheses (e.g. Veeramah et al. 2011; Eriksson et al. 2012; Posth et al. 2016; Ray et al. 2009).

Archaeology and more generally cultural evolution have adopted this modelling paradigm and generative inference frameworks enjoy increasing popularity (e.g. Crema et al. 2014, 2016; Kandler and Laland 2013; Kandler and Shennan 2015; Porčić and Nikolić 2016; Edinborough et al. 2015; Kovacevic et al. 2015; Rubio-Campillo 2016; Thouzeau et al. 2017). They have been applied to questions as diverse as the inference of processes of cultural transmission (e.g. Crema et al. 2014, 2016; Kandler and Shennan 2015) and the estimation of growth rates and population sizes from cultural frequency data (e.g. Porčić and Nikolić 2016), the investigation of the evolution of combat (Rubio-Campillo 2016) or of the coevolution between genes and languages at a regional scale (e.g. Thouzeau et al. 2017).

In this section we briefly introduce the idea of generative inference frameworks (for a detailed discussion see, e.g., Kandler and Powell 2018) and illustrate the applicability of this approach to archaeological data. Thereby we assume that the data provide a description of the dynamic of cultural change by recording the composition of samples of cultural variants (usually drawn from populations of unknown size) at several points in time.

Generative Inference Framework

The generative inference procedure consists of two main steps. The first step comprises the development of a generative model to produce pseudo-data, in our case population-level frequencies of different variants at different points in time conditioned on an assumed cultural transmission process. The second step uses Bayesian techniques such as approximate Bayesian computation to statistically compare theoretical predictions and empirical observations and to derive conclusions about which (mixtures of) transmission processes are consistent with the observable frequency data (and which are not). The outcome of this approach is not only the identification of the most likely underlying transmission process given the empirical data but a description of the breadth of processes that could have produced the these data equally well which in turn can be interpreted as an informal measure of the level of equifinality (Kandler and Powell 2018).

In more detail, the **generative model** is aimed at capturing the main cultural and demographic dynamics of the cultural system considered. Importantly, the model produces data of the same type as the observed data but conditioned on the assumed cultural transmission process. Thereby different transmission processes are expressed by different model parameterisations; the model parameters are denoted by $\theta = (\theta_1, \ldots, \theta_s)$ in the following. In other words, the generative model establishes an explicit causal relationship between the assumed processes of cultural transmission defined by θ and observable population-level patterns of cultural change. There are no restrictions on the type of generative model used. Models ranging from systems of partial differential equations to agent-based simulations have been applied successfully.

As the generative model generates frequency distributions at different points in time it can be designed as a non-equilibrium framework where some of the issues mentioned in section "Equilibrium Assumption" are accounted for. In other words, starting from an initial condition we are interested in the frequencies of the cultural variants at a specific time point and not necessarily at steady state. Naturally this also allows for the incorporation of temporally changing model parameters such

as population size and innovation rate. Consequently, the risk of misinterpreting non-equilibrium dynamics as evidence for the presence or absence of particular cultural transmission processes can be reduced but this modelling choice requires relatively accurate knowledge about the time points at which the observed frequencies are recorded.

Summarising, the generative model produces pseudo-data of the same type as the observed data conditioned on a specific hypothesis of cultural transmission which is parameterised by the values of the model parameters $\theta = (\theta_1, \ldots, \theta_s)$. **Statistical comparisons** then determine which of the cultural transmission processes considered in the generative model could produce pseudo-data similar to the observed data. For that one would ideally determine the likelihood function of the generative model. However, in many cases the likelihood functions cannot be determined easily. **Approximate Bayesian computation (ABC)** (Beaumont et al. 2002; Pritchard et al. 1999) was developed to circumvent this difficulty. In the following we briefly describe the ABC procedure in its simplest form, the rejection algorithm, but note that many extensions have been published (e.g. Beaumont et al. 2002; Blum and François 2010; Marjoram et al. 2003; Toni et al. 2009).

Given observed data D, the ABC approach approximates the joint posterior distribution of the model parameters θ, denoted by $P(\theta|D)$. It does this through repeatedly simulating data $D^\star$ under a generative model with parameter values drawn from their prior distributions $P(\theta)$. These prior distributions describe the possible values the parameter can assume or summarise all prior knowledge researchers may have. Retaining those parameter sets that generate data $D^\star$ sufficiently "close" to the observed data D, and rejecting the rest, results in a random sample from the distribution $P(\theta|d(D, D^\star) \leq \varepsilon)$, where $d(\cdot, \cdot)$ is a distance metric between the observed and simulated data, and ε is a tolerance level determining the approximation to the true posterior $P(\theta|D)$. In situation where the observed data D is high-dimensional it may become a challenge to determine how the distance between empirical and theoretical data should be calculated. Here often summary statistics S (such as the level of cultural diversity) are used which implies that not the distribution $P(\theta|d(D, D^\star) \leq \varepsilon)$ but $P(\theta|d(S, S^\star) \leq \varepsilon)$ is sampled. The choice of appropriate summary statistics is not straightforward and can have crucial consequences on the inference accuracy (the choice of summary statistics is currently an active area of statistical research, e.g., Harrison and Baker 2017). However, those problems are evaded if the distance is calculated based on the 'raw' data D and $D^\star$, in our case by calculating the euclidean distance between the observed and theoretical frequencies.

The output of any ABC procedure is the joint posterior distribution of the model parameters $\theta = (\theta_1, \ldots, \theta_s)$ (and derived from that the marginal posterior distributions for each individual parameter), indicating the range of the parameter space that is able to produce frequency data within a given tolerance level ε of the observed data, and consequently the transmission that is consistent with the data. Therefore the widths of these distributions may provide an informal measure of the level of equifinality. If the posterior distributions are narrow (compared to their corresponding prior distributions), then only a small region of the parameter space is consistent with the data and therefore a large number of transmission processes are not able to produce the observed frequency changes. In this case the data carries a relatively strong signature of the underlying processes of cultural transmission. In contrast, if the distributions are wide, a large region of the parameter space is consistent with the data and therefore many processes of cultural transmission are able to generate very similar population-level frequency patterns (Kandler and Powell 2018).

In practice, performing ABC analyses has been made relatively straightforward by the release of software such as *DIY-ABC* (Cornuet et al. 2008), *ABCtoolbox* (Wegmann et al. 2010), and *R* packages such as *abc* (Csilléry et al. 2012), *abctools* (Nunes and Prangle 2015) and *EasyABC* (Jabot et al. 2013).

Cautionary Notes

One of the main advantages of the generative inference approach described above is that it allows for the statistical evaluation of complex cultural and demographic scenarios. Nevertheless the interpretation of the analysis results has to be done cautiously; it is, after all, an analysis based on an underlying model of cultural change. If this model does not capture the main cultural and demographic processes contributing to the observed temporal frequency changes, the inferences obtained will likely be misleading.

The ABC procedure *always* outputs a joint posterior distribution of the inferred model parameter. But as mentioned before this distribution only approximates the 'true' distribution if the achieved error level ε, i.e. the distance between theoretical and empirical data, is close to zero. If the obtained ε is large and cannot be improved upon the inferred parameter spaces are likely not meaningful and should not be interpreted. This situation may point to an inadequacy of the model, and therefore the assumed processes of cultural transmission, to explain the data. The explanatory value of the obtained posterior distribution can be further investigated by posterior predictive checks (Gelman et al. 2013, see Crema et al. 2016 for an archaeological application). In our situation, these checks consist of sampling values of the model parameters from the joint posterior distribution, inserting these into the generative model and producing theoretical frequencies at the required time points. Repeating this procedure generated theoretical expectations of the frequency ranges for each individual variant based on the joint posterior distribution. The comparison of the observed frequencies of each variant with these frequency ranges allowed the explanatory power of the derived posterior distribution to be assessed. If observations are outside the theoretical expectations, then the inferred cultural transmission processes cannot replicate all aspects of the dynamic of cultural change, indicating a mismatch between theory and data. Additionally, crossvalidation tests or coverage plots have been developed to further investigate the accuracy of the results of the ABC analysis (Csilléry et al. 2012; Prangle et al. 2014; van der Vaart et al. 2015).

A common criticism of the ABC method focuses on the choice of prior distributions which may only be "guessed based upon the subjective opinion of the investigators" (Templeton 2010). This argument is connected to classical objections of Bayesian approaches (Beaumont et al. 2010). The prior distribution represents prior beliefs about the distribution of the parameters and should if possible be defined based on known properties of the system studied (e.g. the duration of archaeological phases, the hypothesised number of social learners, etc.) but may for practical applications necessitate an educated guess (Sunnåker et al. 2013). While informative priors can influence the widths of the posterior distributions, uninformative and flat priors can still yield reasonable parameter estimates. However, Bayes factors (for model comparisons) are highly sensitive to the prior distribution of parameters (Sunnåker et al. 2013).

As already mentioned, the accuracy of the inference depends partly on the way the distance between theoretical and empirical data is calculated. Calculating the difference based on an insufficient summary statistics S instead of the full data D results in discarding likely useful information and the posterior distribution will not be equal to that computed with the full data (Csilléry et al. 2010). Therefore, if possible, inferences should be based on the full data and if this is not possible the summary statistic (or a set of) needs to be carefully chosen (see Csilléry et al. 2010, for a review of strategies of how to do so).

Example

In the following we demonstrate how such an inference framework can be constructed and used by summarising the analysis of an archaeological data set describing the culture of the first farmers in

Fig. 5.5 In each time step, ν variants are generated (For the sake of simplicity we kept ν constant in each time step, however, this assumption is relaxed in the non-equilibrium versions of the model described below.) and the population-level frequencies of the different variants (indicated by different colours) at the end of each phase i describe the accumulation of all η_i production events. It holds $N_i = \nu \eta_i$
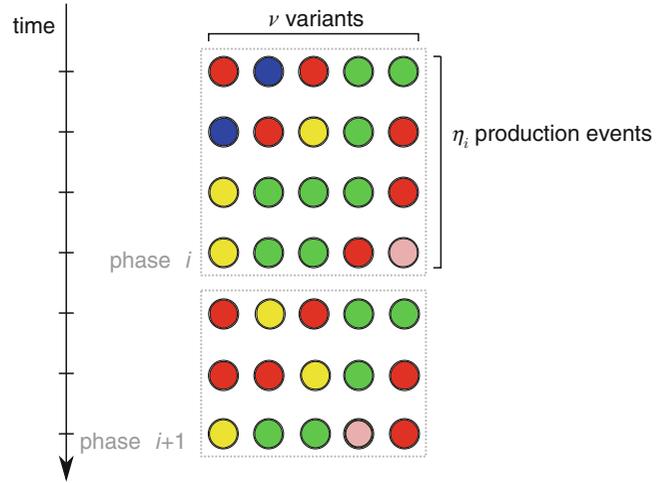
Central Europe in the valley of the Merzbach stream in western Germany, the so-called Linearband-keramik (LBK) culture, from ca. 5300 to 4850 cal. B.C. (see Crema et al. 2016, for the complete analysis). The data set records the frequencies of 36 types of pottery vessels in eight different phases. The aim of this study was to explore whether observed frequency changes in different types of pottery between the different phases are consistent with a specific hypothesis about the underlying cultural transmission process, in particular unbiased transmission and frequency-dependent transmission, under the assumptions that (1) the cultural system producing those frequencies is at equilibrium and (2) certain aspects of the system change over time.

First, we developed the generative model. As mentioned above, the model had to generate frequency changes of the different cultural variants between two successive phase conditioned on a specific process of cultural transmission but also reflect the essential features of the way the frequencies have been generated. Crucially this model had to take into account that the observed frequencies in phase i describe the composition of a sample of size n_i but not of the population of unknown size. Any evolutionary process, however, acts on the whole population and not only on the sample and consequently processes of cultural transmission need to be modelled on the population-level. To do so we assumed that the population size N_i can be determined from the sample size through the relation $N_i = n_i/r$ with $0 < r < 1$. The variable r describes the expected fraction of the population that has been sampled. We further assumed that the N_i cultural variants are the result of the accumulation of η_i production events of ν_t variants throughout phase i (see Fig. 5.5). In other words, it holds $N_i = \sum_{t=t_i}^{t_i+\eta_i-1} \nu_t$.

The composition of the ν_t variants in each production event is determined by the assumed process of cultural transmission which guides the decision to add an instance of a certain variant based on the available social information. Here social information consists of a sampling pool, composed of the cultural variants of the last w production events. In detail, an instance of variant j, $j = 1, \ldots, k$ is chosen to be added to the population according to the probability

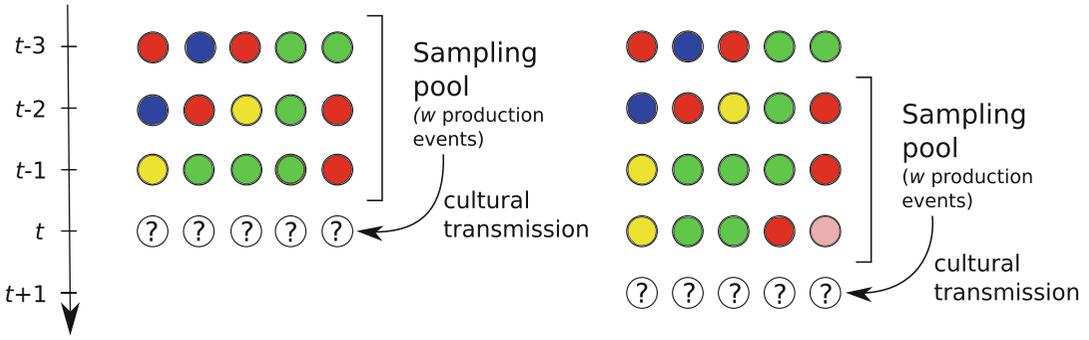$$\pi_j = \frac{p_j^{1-b}}{\sum_{l=1}^{k} p_l^{1-b}} (1 - \mu) \tag{5.13}$$

Fig. 5.6 Scheme of the transmission dynamic. In each production event ν instances of cultural variants are produced and the probability with which one of the existing variants is produced is given by Eq. (5.13). For instance, in the left figure any of the five instances at time t turns red with probability $\frac{(5/15)^{1-b}}{(5/15)^{1-b}+(2/15)^{1-b}+(6/15)^{1-b}+(2/15)^{1-b}}(1-\mu)$. With probability μ an entirely new variant is introduced (see the pink variant at time $t+1$ in the right figure)

where p_i describes the relative frequency of variant i in the sampling pool, the parameter b controls the strength of frequency-dependent transmission, and μ stands for the innovation rate (see Fig. 5.6). Importantly, setting $b = 0$ reduces Eq. (5.13) to unbiased transmission (cf. Eq. (5.1)).

This process is repeated until η_i production events have produced in total N_i instances of cultural variants. In this way we have generated a theoretical population of variants conditioned on a specific process of cultural transmission (parameterised by the value of parameters b, μ, and w) given by Eq. (5.13). In order to create data that can be compared to the observed data we lastly drew n_i variants randomly from this population. Summarising, we developed a model that, based on certain cultural and demographic assumptions, can describe the dynamic of cultural change conditioned on a specific process of cultural transmission. Further, this model is flexible enough to replicate equilibrium and non-equilibrium systems. In order to explore the consequences of these modelling choices we analysed the following three versions of the model (see Crema et al. 2016, for details).

Equilibrium Condition To generate data for a system at equilibrium we repeated the above described dynamic sufficiently long (starting from an arbitrary initial condition) under constant parameter values, i.e. $\theta = [\mu, b, \nu, r, w]$ did not change over time. This means that both, the process of cultural transmission and the number of variants ν generated in each production event were kept constant. After this burn-in period, we generated populations of sizes N_i for all seven consecutive phases i and randomly sampled from these populations n_i instances of cultural variants. If the observed data was generated by a process of unbiased transmission (i.e. $b = 0$), the posterior distribution of b should be centred around 0.

Variable Population Condition We relaxed the assumption that the number of cultural variants generated in each production events is constant over time but still assumed that the process of cultural transmission is the same throughout all phases. Therefore the main differences to the equilibrium version are that (1) ν is time-dependent, and hence its value can change at each production event; and (2) the model is no longer initialised through a burn-in phase. Instead the initial sampling pool is derived from the observed frequencies at the end of phase $i - 1$ using the Dirichlet distribution approach (Gelman et al. 2013).

Variable Population-Transmission Process Condition We allowed both, the number of cultural variants produced per production event and the cultural transmission process to vary over time. For this we applied the same routine described in the variable population size condition but considered each phase separately. Thus we generated a sampling pool at the beginning of each phase i (following the Dirichlet approach and using the observed frequencies at the end of phase $i - 1$), let η_i production

events occur to generate a population of cultural variants at the end of phase i conditioned on the cultural transmission process defined by the parameters b_i, μ_i, and w_i and drew a random sample of size n_i.

Now to infer which cultural transmission hypotheses are consistent with observed data we applied the ABC rejection procedure. For each version of the model we determined the parameter values which were able to generate variant frequencies "close" to the observed ones. As distance measure we used the euclidean distance between theoretical and observed frequencies. In detail, we sampled parameter combinations from prior distributions of b, μ, w, r, and ν, executed the simulation model, and measured the euclidean distance ε between the obtained theoretical data and observed data. From the pool of s iterations of this procedure a proportion α, representing the lowest values of the error level ε, are retained and the parameter values associated with these were assessed. We set $s = 10^7$ and $\alpha = 2 \times 10^{-5}$ for all three versions of the model. Sample codes of the simulation model, the ABC framework, and the data set can be downloaded at https://github.com/ercrema/CulturalTransmissionModel.

The output of this inference procedure is the joint posterior distribution of the inferred parameters $\theta = [b, \mu, w, r, \nu]$. This distribution reveals which parameter values, and therefore which processes (amongst the models considered in the generative model) could replicate the observed samples.

From the range of the corresponding marginal distributions we concluded that the strength of frequency-dependent transmission b was by far the most informative parameter, pointing to a relationship between processes of cultural transmission and observable patterns of cultural change. The posterior distributions of the other model parameters covered in all cases nearly the same range as their respective prior distributions indicating that the data is not informative about those parameters.

The 95% highest posterior density interval (HPDI) of the posterior distribution of b for all three model versions are shown in Fig. 5.7. But before exploring the differences between the versions we take a closer look at the distributions. The HPDIs describe the parameter ranges that could have generated the observed frequencies. In all cases we *cannot* infer a single transmission processes (such as unbiased transmission) as the sole candidate for explaining the data. We instead observed an equifinal situation: a range of values of b (and therefore frequency-dependent transmission of various strengths) are consistent with the data. The median values (indicated by the black horizontal lines) point to the most likely process but nevertheless all other values of b within HDPI have produced frequency data with a similar error tolerance ε.

In the equilibrium version, the median value of the distribution of b is 0.028 which would suggest the presence of some degree of weak negative frequency-dependent transmission, albeit the HPDI covers a range between -0.005 and 0.102 (see Fig. 5.7). The variable population version appeared to show a stronger support of negative frequency-dependent transmission, with the entire 95% HPDI (0.015–0.134) larger than zero and a median value of 0.066. To further analyse the data we performed a posterior predictive check of the equilibrium and the variable population versions where both versions showed a relatively poor performance with a considerable number of observations outside their expected frequency ranges. This suggested that neither equilibrium conditions nor a changing number of copying events can explain the observed data.

In the variable population-transmission process version the posterior distributions of b for the individual phases suggest fluctuations (although the 95% HPDI of b is not sufficiently narrow to fully dismiss competing models), with earlier stages showing stronger support for negative frequency-dependent transmission and intermediate phases for a positive frequency-dependent transmission (see Fig. 5.7). The posterior predictive check of the variable population-transmission mode version showed a clear improvement with the 95% range of the model prediction including all observed frequencies except three variant types suggesting a change in the process of cultural transmission between successive phases. Consequently, this kind of analysis may also reveal individual variants which are "special" in the sense that they do not follow the general dynamic of the system and

Fig. 5.7 Marginal posterior distributions of the strength of frequency-dependent transmission *b* for the equilibrium (across all phases), variable population (across all phases), and variable population-transmission (per individual phase) version of the generative model
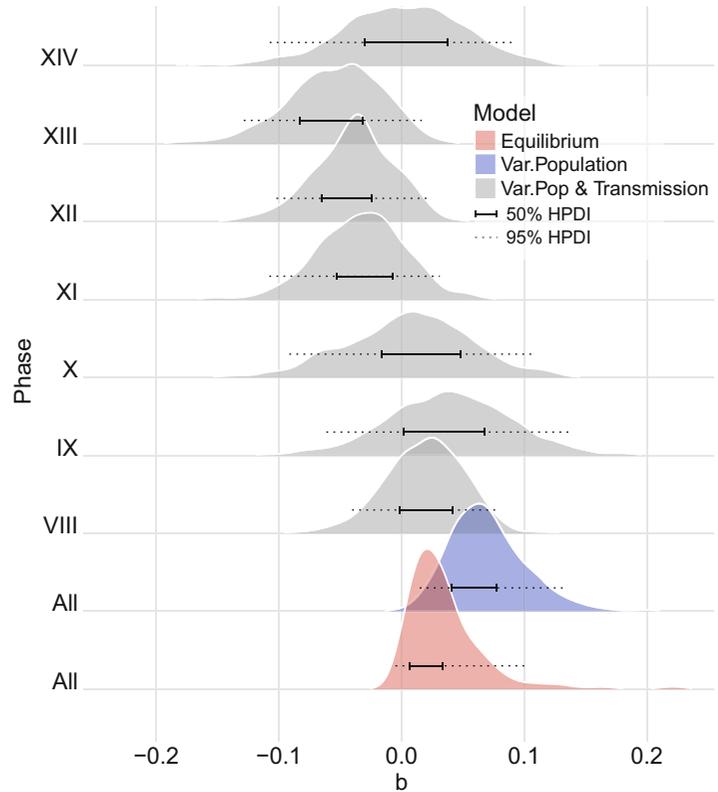
potentially are selected for or against by other forms of cultural transmission such as content biased transmission.

Summarising, our attempts at inferring patterns of cultural transmission in the Merzbach assemblage revealed a cultural system that is unlikely to be in equilibrium conditions. Instead our results hinted at the possibility of shifts between negative and positive frequency-dependent transmission. But crucially the widths of the obtained posterior distributions showed that there exist limits to the inference of underlying processes from cultural transmission from the observed population-level frequency data.

Discussion and Conclusion

Neutral theory has been central to applying quantitative evolutionary thinking to archaeology. It suggests that temporal changes in cultural assemblages can be explained by a process of unbiased transmission, random drift, and innovation and the mathematical frameworks developed in population genetics provided means to analyse the patterns of assemblage variation by exploring their consistency (or inconsistency) with neutral expectations. In other words, these frameworks present a way of testing whether observed population-level data describing the frequencies of various cultural variants can be explained by the process of unbiased transmission. Archaeological applications of neutral theory have mainly focused on ceramic assemblages and the comparison between expected and observed levels of cultural diversity (e.g. Neiman 1995; Shennan and Wilkinson 2001; Lipo 2001; Kohler et al.

2004; Schauer 2008; Steele et al. 2010; Premo and Scholnick 2011). Most of these studies concluded consistency between neutral theory and the observed assemblages.

But the assumption of unbiased transmission is often at odds with what archaeologist and anthropologists know about social systems. So what can we learn from the consistency of between neutral theory and empirical data? It is worth mentioning that consistency has not been interpreted as evidence that all individuals adopt cultural variants at random. Consistency does mean that no selective processes have to be invoked in order to explain the observed population-level frequency patterns making neutral theory the most parsimonious mechanism to account for the commonly observed life cycles of artefact attributes, such as decoration patterns: initial appearance, rise and decline in popularity followed by eventual extinction. But we argue in this chapter that we need to be careful with interpreting the consistency (or inconsistency) of archaeological data with theoretical neutral expectations as evidence for the absence (or presence) of selective force in the population. While neutral theory does very well in replicating population-level frequency patterns, many alternative processes of cultural transmission may do so, too, making the pattern-matching approach a fairly weak approach for inferring underlying transmission processes (see, e.g., Gotelli and McGill 2006; Rosindell et al. 2012, for discussions of this in the ecological context). Now one may argue that this is not a problem of neutral theory but of the (sparse) data, or the test statistics used. In fact, stronger tests of departures from neutrality, e.g. involving analyses of the temporal dynamic of cultural change, have proven more powerful in distinguishing between neutral evolution and selective forces (e.g. Sindi and Dale 2016). But in archaeological applications it is unlikely to easily improve the temporal and/or spatial resolution of the data and therefore the pattern-process issue has to be accounted for.

Here we advocated the use of a generative inference approach as one possibility to go beyond the hypothesis testing approach. At the heart of this framework is a generative model, which captures the main cultural and demographic properties of the system considered. These models establish a causal link between model parameters controlling the strengths of underlying evolutionary processes and observable population-level patterns; in our case between parameters controlling the strengths and nature of the cultural transmission processes considered and population-level frequencies of cultural variants. Bayesian inference techniques, such as ABC, can then evaluate whether a specific process of cultural transmission is able to produce frequency patterns consistent with the observed ones. The main advantage of this approach that it allows for relatively complex, potentially non-equilibrium generative models which may include all information researches may have about the system considered (e.g. time-averaging processes, changes in population size, precise time points of observations). This comes at the cost that the statistical comparison is computationally very intensive, however, modern-day computing equipment (especially the use of high-performance clusters) can overcome this obstacle.

The outcome of this inference approach are posterior distributions of the model parameters, specifying the transmission processes, that are consistent with the observed frequency data. While there are a number of important factors, potentially influencing the accuracy of the analysis, to consider (see, e.g., Kandler and Powell 2018, for more details), the widths of the posterior distributions may be indicative of the amount of information about the underlying transmission processes contained in the data. Narrow posterior distributions indicate that the data carries a relatively strong signature of these processes, while wider distributions suggest that the data is largely uninformative or that the models considered do not provide an adequate description of the cultural system. Therefore this approach does not only allow for the identification of the most likely underlying learning process given the empirical data but for a description of the breadth of processes that could have produced these data equally well, providing an informal measure of equifinality.

Applications of this framework to archaeological data have shown that researchers should not expect to be able to infer a single cultural transmission process that solely can explain the data. Different transmission processes will be consistent with the data. And it is one of the strengths of mathematical modelling approaches to archaeological data to help us understand these limits

to inferring underlying transmission processes from population-level frequency data and therefore to identify which kinds of questions can be answered with which kinds of data. The value of the analyses similar to the one described in section "Application to Archaeology" lies in the exclusion of transmission processes that could not have produced the observed data and therefore in a reduction of the pool of potential hypotheses. Subsequently, different lines of evidence may be used to reduce this pool even further.

References

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

Baum, W., Richerson, P., Efferson, C., & Paciotti, B. (2004). Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. *Evolution and Human Behavior, 25*(5), 305–326.

Beaumont, M., Zhang, W., & Balding, D. (2002). Approximate Bayesian computation in population genetics. *Genetics, 162*(4), 2025–2035.

Beaumont, M. A., Nielsen, R., Robert, C., Hey, J., Gaggiotti, O., Knowles, L., et al. (2010). In defence of model-based inference in phylogeography. *Molecular Ecology, 19*(3), 436–446.

Beheim, B., Thigpen, C., & McElreath, R. (2014). Strategic social learning and the population dynamics of human behavior: The game of go. *Evolution and Human Behavior, 35*(5), 351–357.

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

Bentley, R., Hahn, M., & S. Shennan (2004). Random drift and culture change. *Proceedings of the Royal Society of London B: Biological Sciences, 271*(1547), 1443–1450.

Bentley, R., Lipo, C., Herzog, H., & Hahn, M. (2007). Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior, 28*(3), 151–158.

Binford, L. (1963). "Red Ocher" caches from the Michigan area: A possible case of cultural drift. *Southwestern Journal of Anthropology, 19*(1), 89–108.

Blum, M., & François, O. (2010). Non-linear regression models for approximate Bayesian computation. *Statistics and Computing, 20*(1), 63–73.

Bortolini, E. (2014). *An evolutionary and quantitative analysis of construction variation in prehistoric monumental burials of eastern Arabia*. PhD thesis, UCL.

Boyd, R., & Richerson, P. (1985). *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.

Brantingham, P. J., & Perreault, C. (2010). Detecting the effects of selection and stochastic forces in archaeological assemblages. *Journal of Archaeological Science, 37*(12), 3211–3225.

Caldwell, C., & Millen, A. (2009). Social learning mechanisms and cumulative cultural evolution: Is imitation necessary? *Psychological Science, 20*(12), 1478–1483.

Cavalli-Sforza, L., & Feldman, M. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton, NJ: Princeton University Press.

Chisholm, R., & O'Dwyer, J. (2014). Species ages in neutral biodiversity models. *Theoretical Population Biology, 93*, 85–94.

Cochrane, E. (2009). Evolutionary explanation and the record of interest: Using evolutionary archaeology and dual-inheritance theory to explain the archaeological record. In S. Shennan (Ed.), *Pattern and process in cultural evolution* (pp. 113–132). Berkeley, CA: University of California Press.

Cornuet, J.-M., Santos, F., Beaumont, M., Robert, C., Marin, J.-M., Balding, D., et al. (2008). Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics, 24*(23), 2713.

Coultas, J. (2004). When in rome ... an evolutionary perspective on conformity. *Group Processes & Intergroup Relations, 7*(4), 317–331.

Crema, E. (2018). Statistical inference and archaeological simulations. *The SAA Archaeological Record, 18*(2), 20–23.

Crema, E., Edinborough, K., Kerig, T., & Shennan, S. (2014). An approximate Bayesian computation approach for inferring patterns of cultural evolutionary change. *Journal of Archaeological Science, 50*, 160–170.

Crema, E., Kandler, A., & Shennan, S. (2016). Revealing patterns of cultural transmission from frequency data: Equilibrium and non-equilibrium assumptions. *Scientific reports*, 6.

Csilléry, K., Blum, M., Gaggiotti, O., & François, O. (2010). Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution, 25*(7), 410–418.

Csilléry, K., François, O., & Blum, M. (2012). ABC: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution, 3*(3), 475–479.

Dunnell, R. (1978). Style and function: A fundamental dichotomy. *American Antiquity, 43*(2), 192–202.

Edinborough, K., Shennan, S., Crema, E., & Kerig, T. (2015). An ABC of lithic arrowheads: A case study from south-eastern France. In *Neolithic diversities. Acta Archaeologica Lundensia, series 8o* (Vol. 65, pp. 213–223). Lund: Department of Archaeology and Ancient History, Lund University.

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

ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R., & Baronchelli, A. (2017). Evolutionary dynamics of the cryptocurrency market. *Royal Society Open Science, 4*(11), 170623.

Eriksson, A., Betti, L., Friend, A., Lycett, S., Singarayer, J., von Cramon-Taubadel, N., et al. (2012). Late pleistocene climate change and the global expansion of anatomically modern humans. *Proceedings of the National Academy of Sciences, 109*(40), 16089–16094.

Evans, T., & Giometto, A. (2011) Turnover rate of popularity charts in neutral models. *arXiv preprint. arXiv:1105.4044*.

Ewens, W. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology, 3*(1), 87–112.

Ewens, W. (2004). *Mathematical population genetics 1: Theoretical introduction* (Vol. 27). New York, NY: Springer.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gotelli, N., & McGill, B. (2006). Null versus neutral models: What's the difference? *Ecography, 29*(5), 793–800.

Hahn, M., & Bentley, R. (2003). Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society of London B: Biological Sciences, 270*(Suppl 1), S120–S123.

Harrison, J., & Baker, R. (2017). An automatic adaptive method to combine summary statistics in approximate Bayesian computation. *arXiv: 1703.02341v1*.

Henrich, J. (2001). Cultural transmission and the diffusion of innovations: Adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. *American Anthropologist, 103*(4), 992–1013.

Henrich, J., & Boyd, R. (2002). Cognition and culture: Why cultural evolution does not require replication of representations. *Journal of Cognition and Culture, 2*, 87–112.

Henrich, J., & Broesch, J. (2011). On the nature of cultural transmission networks: Evidence from Fijian villages for adaptive learning biases. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1567), 1139–1148.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews, 12*(3), 123–135.

Herzog, H., Bentley, R., Hahn, M. (2004). Random drift and large shifts in popularity of dog breeds. *Proceedings of the Royal Society of London B: Biological Sciences, 271*(Suppl 5), S353–S356.

Hoppitt, W., Boogert, N., & Laland, K. (2010). Detecting social transmission in networks. *Journal of Theoretical Biology, 263*(4), 544–555.

Hubbell, S. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton, NJ: Princeton University Press.

Jabot, F., Faure, T., & Dumoulin, N. (2013). Easyabc: Performing efficient approximate Bayesian computation sampling schemes using R. *Methods in Ecology and Evolution, 4*(7), 684–687.

Kandler, A., & Laland, K. (2013). Tradeoffs between the strength of conformity and number of conformists in variable environments. *Journal of Theoretical Biology 332*, 191–202.

Kandler, A., & Powell, A. (2018). Generative inference in cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*, 20170056.

Kandler, A., & Shennan, S. (2013). A non-equilibrium neutral model for analysing cultural change. *Journal of Theoretical Biology, 330*, 18–25.

Kandler, A., & Shennan, S. (2015). A generative inference framework for analysing patterns of cultural change in sparse population data with evidence for fashion trends in LBK culture. *Journal of The Royal Society Interface, 12*(113), 20150905.

Kandler, A., Wilder, B., & Fortunato, L. (2017). Inferring individual-level processes from population-level patterns in cultural evolution. *Open Science, 4*(9), 170949.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature, 217*(5129), 624–626.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.

Kimura, M. (1991). The neutral theory of molecular evolution: A review of recent evidence. *The Japanese Journal of Genetics, 66*(4), 367–386.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681–10686.

Kohler, T., VanBuskirk, S., & Ruscavage-Barz, S. (2004). Vessels and villages: Evidence for conformist transmission in early village aggregations on the Pajarito Plateau, New Mexico. *Journal of Anthropological Archaeology, 23*(1), 100–118.

Kovacevic, M., Shennan, S., Vanhaeren, M., d'Errico, F., & Thomas, M. (2015). Simulating geographical variation in material culture: Were early modern humans in Europe ethnically structured? In A. Mesoudi, & K. Aoki (Eds.), *Learning strategies and cultural evolution during the Palaeolithic* (pp. 103–120). Berlin: Springer.

Laland, K. (2004). Social learning strategies. *Animal Learning & Behavior, 32*(1), 4–14.

Leigh, E. (2007). Neutral theory: A historical perspective. *Journal of Evolutionary Biology, 20*, 2075–2091.

Lipo, C. (2001). *Science, style and the study of community structure: An example from the central Mississippi River valley* (Vol. 918). Oxford: British Archaeological Reports Limited.

Lipo, C. P., Madsen, M. E., Dunnell, R. C., & Hunt, T. (1997). Population structure, cultural transmission, and frequency seriation. *Journal of Anthropological Archaeology, 16*(4), 301–333.

Lycett, S. (2015). Cultural evolutionary approaches to artifact variation over time and space: Basis, progress, and prospects. *Journal of Archaeological Science, 56*, 21–31.

Madsen, M. (2012). Unbiased cultural transmission in time-averaged archaeological assemblages. *ArXiv e-prints*.

Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences, 100*(26), 15324–15328.

McElreath, R., Bell, A., Efferson, C., Lubell, M., Richerson, P., & Waring, T. (2008). Beyond existence and aiming outside the laboratory: Estimating frequency-dependent and pay-off-biased social learning strategies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1509), 3515–3528.

McElreath, R., Lubell, M., Richerson, P., Waring, T., Baum, W., Edsten, E., et al. (2005). Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior, 26*(6), 483–508.

Mesoudi, A., & O'Brien, M. (2008). The cultural transmission of great basin projectile-point technology I: An experimental simulation. *American Antiquity, 73*(1), 3–28.

Moran, P. (1958). Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 54, pp. 60–71). Cambridge: Cambridge University Press.

Morgan, T., Rendell, L., Ehn, M., Hoppitt, W., & Laland, K. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society of London B: Biological Sciences, 279*(1729), 653–662.

Neiman, F. (1995). Stylistic variation in evolutionary perspective: Inferences from decorative diversity and interassemblage distance in Illinois woodland ceramic assemblages. *American Antiquity, 60*, 7–36.

Nunes, M., & Prangle, D. (2015). abctools: An R package for tuning approximate Bayesian computation analyses. *The R Journal, 7*(2), 189–205.

O'Brien, M., Lyman, R., Mesoudi, A., & Van Pool, T. (2010). Cultural traits as units of analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*, 3797–3806.

O'Dwyer, J., & Kandler, A. (2017). Inferring processes of cultural transmission: The critical role of rare variants in distinguishing neutrality from novelty biases. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1735), 20160426.

O'Dwyer, J., Sharpton, T., & Kembel, S. (2015). Backbones of evolutionary history test biodiversity theory in microbial communities. *Proceedings of the National Academy of Science, 112*, 8356–8361.

Perreault, C. (2018). Time-averaging slows down rates of change in the archaeological record. *Journal of Archaeological Method and Theory, 25*, 953–964.

Pocklington, R., & Best, M. (1997). Cultural evolution and units of selection in replicating text. *Journal of Theoretical Biology, 188*, 79–87.

Porčić, M., & Nikolić, M. (2016). The approximate Bayesian computation approach to reconstructing population dynamics and size from settlement data: Demography of the Mesolithic-Neolithic transition at Lepenski Vir. *Archaeological and Anthropological Sciences, 8*(1), 169–186.

Posth, C., Renaud, G., Mittnik, A., Drucker, D., Rougier, H., Cupillard, C., et al. (2016). Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a late glacial population turnover in Europe. *Current Biology, 26*(6), 827–833.

Prangle, D., Blum, M., Popovic, G., & Sisson, S. (2014). Diagnostic tools for approximate Bayesian computation using the coverage property. *Australian & New Zealand Journal of Statistics, 56*(4), 309–329.

Premo, L. (2010). Equifinality and explanation: The role of agent-based modeling in postpositivist archaeology. *Simulating change: Archaeology into the twenty-first century* (pp. 28–37). Salt Lake City: University of Utah Press.

Premo, L. (2014). Cultural transmission and diversity in time-averaged assemblages. *Current Anthropology, 55*(1), 105–114.

Premo, L., & Scholnick, J. (2011). The spatial scale of social learning affects cultural diversity. *American Antiquity, 76*(1), 163–176.

Pritchard, J., Seielstad, M., Perez-Lezaun, A., & Feldman, M. (1999). Population growth of human y chromosomes: A study of y chromosome microsatellites. *Molecular Biology and Evolution, 16*, 1791–1798.

Ramsey, G., & De Block, A. (2017). Cognitive culture: Theoretical and empirical insights into social learning strategies. *The British Journal for the Philosophy of Science, 68*(2), 305–328.

Ray, N., Wegmann, D., Fagundes, N., Wang, S., Ruiz-Linares, A., & Excoffier, L. (2009). A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Molecular Biology and Evolution, 27*(2), 337–345.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M., et al. (2010). Why copy others? Insights from the social learning strategies tournament. *Science, 328*(5975), 208–213.

Richerson, P., & Boyd, R. (2008). Response to our critics. *Biology & Philosophy, 23*(2), 301–315.

Rorabaugh, A. N. (2014). Impacts of drift and population bottlenecks on the cultural transmission of a neutral continuous trait: An agent based model. *Journal of Archaeological Science, 49*, 255–264.

Rosindell, J., Hubbell, S., He, F., Harmon, L., & Etienne, R. (2012). The case for ecological neutral theory. *Trends in Ecology & Evolution, 27*(4), 203–208.

Rubio-Campillo, X. (2016). Model selection in historical research using approximate Bayesian computation. *PLoS One, 11*(1), e0146491.

Schauer, P. (2008). *Cultural evolution in the age of Athens: Drift and selection in Greek figure-painted pottery.* PhD thesis, UCL.

Scholnick, J. (2010). *Apprenticeship, cultural transmission and the evolution of cultural traditions in historic New England gravestones.* PhD thesis, University of Arizona.

Shennan, S. (2011). Descent with modification and the archaeological record. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 366*(1567), 1070–1079.

Shennan, S., & Wilkinson, J. (2001). Ceramic style change and neutral evolution: A case study from Neolithic Europe. *American Antiquity, 66*, 577–593.

Sindi, S., & Dale, R. (2016). Culturomics as a data playground for tests of selection: Mathematical approaches to detecting selection in word use. *Journal of Theoretical Biology, 405*, 140–149.

Slatkin, M. (1994). An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research, 64*(01), 71–74.

Slatkin, M. (1996). A correction to the exact test based on the Ewens sampling distribution. *Genetical Research, 68*(03), 259–260.

Steele, J., Glatz, C., & Kandler, A. (2010). Ceramic diversity, random copying, and tests for selectivity in ceramic production. *Journal of Archaeological Science, 37*(6), 1348–1358.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Computational Biology, 9*(1), e1002803.

Templeton, A. R. (2010). Coherent and incoherent inference in phylogeography and human evolution. *Proceedings of the National Academy of Sciences, 107*(14), 6376–6381.

Thouzeau, V., Mennecier, P., Verdu, P., & Austerlitz, F. (2017). Genetic and linguistic histories in central Asia inferred using approximate Bayesian computations. *Proceedings of the Royal Society B: Biological Sciences, 284*(1861), 20170706.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface, 6*(31), 187–202.

van der Vaart, E., Beaumont, M., Johnston, A., & Sibly, R. (2015). Calibration and evaluation of individual-based models using approximate Bayesian computation. *Ecological Modelling, 312*, 182–190.

Veeramah, K., Wegmann, D., Woerner, A., Mendez, F., Watkins, J., Destro-Bisol, G., et al. (2011). An early divergence of Khoesan ancestors from those of other modern humans is supported by an abc-based analysis of autosomal resequencing data. *Molecular Biology and Evolution, 29*(2), 617–630.

Volkov, I., Banavar, J., Hubbell, S., & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature, 424*(6952), 1035.

von Bertalanffy, L. (1969). *General system theory: Foundations, development, applications* (revised edition). New York, NY: Braziller.

Wang, S., Chen, A., Fang, J., & Pacala, S. (2013). Why abundant tropical tree species are phylogenetically old. *Proceedings of the National Academy of Sciences, 110*, 16039–16043.

Watterson, G. (1977). Heterosis or neutrality? *Genetics, 85*(4), 789–814.

Watterson, G. (1978). The homozygosity test of neutrality. *Genetics, 88*(2), 405–417.

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics, 11*(1), 116.

Wiessner, P. (1983). Style and social information in Kalahari San projectile points. *American Antiquity, 48*(2), 253–276.