

Anne Kandler and Adam Powell

## Abstract

Social learning has been identified as one of the fundamentals of culture and therefore the understanding of why and how individuals use social information presents one of the big questions in cultural evolution. To date much of the theoretical work on social learning has been done in isolation of data. Evolutionary models often provide important insight into which social learning strategies are expected to have evolved but cannot tell us which strategies human populations actually use. In this chapter we explore how much information about the underlying learning strategies can be extracted by analysing the temporal occurrence or usage patterns of different cultural variants in a population. We review the previous methodology that has attempted to infer the underlying social learning processes from such data, showing that they may apply statistical methods with insufficient power to draw reliable inferences. We then introduce a generative inference framework that allows robust inferences on the social learning processes that underlie cultural frequency data. Using developments in population genetics—in the form of generative simulation modelling and approximate Bayesian computation—as our model, we demonstrate the strength of this method with an example based on simulated data.

## Keywords

Social learning • Cultural evolution • Generative inference • Approximate Bayesian computation

## 7.1 Introduction

Much research over the last decades has been devoted to investigating the dynamics of social learning, one of the fundamentals of culture. Social learning, defined as learning facilitated by observations of, or interactions with, another individual or its products (Heyes 1994; Hoppitt and Laland 2013), occurs in various forms and is widespread across the

animal kingdom (Hoppitt and Laland 2013). While it can facilitate the spread of adaptive information, accumulated over generations, throughout a population, it also allows the transmission of outdated, misleading or inappropriate information especially in changing environmental conditions (e.g. Giraldeau et al. 2002). So understanding why and how individuals use social information is one of the big challenges in cultural evolution (Rendell et al. 2010) and the focus of a highly active, interdisciplinary debate. However, much of the modelling work to date has been largely theoretical, in isolation of real data. While models often provide important insights into which social learning strategies are expected to have evolved in changing environmental conditions, they are often unable to make robust inferences about which strategies human populations actually use based on

A. Kandler (✉)

Department of Mathematics, City University London, London, UK  
e-mail: [anne.kandler.1@city.ac.uk](mailto:anne.kandler.1@city.ac.uk)

A. Powell

Palaeogenetics Group, Institute of Anthropology, University of Mainz, Mainz, Germany

observed cultural datasets. Much available empirical data is in the form of frequency distributions of a number of different variants of a cultural trait in the population at a certain point in time (e.g. Bentley et al. 2004) or of time-series that describe the dynamics of the frequency change of cultural variants over time (e.g. Henrich 2001), and often comprise a limited or chronologically-sparse sample from the whole population. Attempting to infer the underlying evolutionary processes from such population-level data poses a classical inverse problem, but despite widespread interest from researchers across archaeology, biological anthropology, animal behaviour and psychology much of the currently available methodology may be limited in its ability to reliably identify which learning strategies (or combinations thereof) could have produced the observed data. While a large number of possible learning strategies have been identified, such as direct-biased social learning, conformist social learning or copy-when-uncertain (e.g. Boyd and Richerson 1985; Laland 2004; Morgan et al. 2012), the inferential frameworks utilised often lack the necessary statistical power to distinguish between the, often very similar, variant frequency distributions generated under those learning strategies. This is especially true when the empirical data is sparse, and it is important to recognise when frequency data lack sufficient information. In such situations inference has to be based on other lines of evidence.

In this chapter we review some of these theoretical inferential approaches, focusing on their application to human learning strategies, and then present a non-equilibrium generative framework which is aimed at robustly addressing the current mismatch between theory and data. This generalizable framework models the frequency change of different variants of a cultural trait present in a population under the influence of various parameterised asocial and social learning strategies, and coupled with a powerful statistical technique called approximate Bayesian computation (ABC) it allows us to delineate the conditions under which such frequency data carry reliable signatures of the underlying learning strategies which generated them. The suggested approach provides a comprehensive and statistically powerful basis for determining possible mixture of learning strategies that are consistent with even sparse observed cultural frequency data, and therefore may result in a substantial reduction of the set of evolutionary hypotheses that could have produced the data.

## 7.2 Previous Research

Research to establish the presence of particular learning strategies in human populations is mainly centred around experimental, laboratory-based and theoretical, modelling-based approaches. Laboratory-based experiments have focused on uncovering the variety and subtlety of human social

learning strategies (Rendell et al. 2011). ‘Microsocieties’ (e.g. Coultas 2004; Baum et al. 2004; McElreath et al. 2008; Morgan et al. 2012) and diffusion chain experiments (e.g. Mesoudi and O’Brien 2008; Caldwell and Millen 2008; Kirby et al. 2008) have proven to provide powerful frameworks for revealing the existence of specific learning strategies in human populations. For a comprehensive review of this experimental literature see (Rendell et al. 2011). In this chapter we focus on theoretical approaches and consider in the following the merits and limitations of the use of adoption curves (Sect. 7.2.1), power-law distributions (Sect. 7.2.2), model selection frameworks (Sect. 7.2.3) and formal population genetic and game-theoretic modelling (Sect. 7.2.4) when investigating the learning strategies employed by human populations.

### 7.2.1 Adoption Curve

Adoption curves detail the cumulative number of individuals that have acquired a specific cultural variant at any given point in time and therefore describe the temporal dynamic of the spread of novel cultural variants through the population. Many researchers suggested that the shape of adoption curves can be used to infer information about the underlying learning strategies employed by human populations; in particular about the presence or absence of social learning. Here mathematical models predicted accelerating adoption curves in situations when social information is used. The acceleration is caused by the frequency-dependent nature of social learning: the more individuals that have acquired the novel cultural variant the higher the probability becomes of others learning it socially, and is often described by a sigmoidal (or *S*-shaped) adoption curve (Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985; Laland et al. 1996). In contrast, asocial learning is predicted to result in decelerating, often *r*-shaped, adoption curves (Boyd and Richerson 1985; Laland et al. 1996). Henrich (2001) argued that asocial learning alone cannot produce *S*-shaped adoption patterns, and therefore that *S*-shaped dynamics must imply the presence of social learning strategies, in various forms, in the population. This reasoning would suggest that the presence or absence of social learning can be inferred directly from the temporal adoption dynamic of a cultural variant, and in fact many episodes of the diffusion of innovations through human populations exhibit an *S*-shaped adoption pattern (Rogers 2003). While it is undoubtedly true that social learning results in accelerating *S*-shaped adoption curves it has been demonstrated that the diagnostic power of the shape of the adoption curve is greatly reduced when allowing for individual heterogeneity or complexity in the adoption process (Reader 2004) or population structure (Franz and Nunn 2009). For example, Rogers (2003) assumed that the adoption time (defined as

the time it takes an individual to adopt the cultural trait) is normally distributed and showed that in this situation asocial learning alone can produce *S*-shaped adoption curves. Therefore, individual differences in the propensity to adopt a cultural trait are likely to influence the adoption dynamic and researchers have to be cautious when attempting to infer the presence or absence of social learning from adoption curve data (Cavalli-Sforza and Feldman 1981). Furthermore, if there is a number of discrete steps required to adopt a cultural variant, the cumulative diffusion curve may be *S*-shaped, even if individuals learn the tasks asocially (Kendal et al. 2007; Hoppitt et al. 2010b). These ambiguities in discriminating between asocial and social learning by analysing the shape of the adoption curve demonstrate that this approach might have limited power when used as an inference tool.

Adoption curves have also been the subject of intense research in marketing science. Bass (1969) proposed that the population of adopters can be divided into innovators and imitators and that the shape of the adoption curve will vary as a function of their relative importance. The Bass model (and its numerous generalisations) has been widely used in forecasting, especially the forecast of new products' sales (e.g. Mahajan et al. 1995). Alternatively to this social learning hypothesis, threshold models have been proposed to explain observed temporal adoption dynamics. Here individuals adopt an innovation as soon as the utility of the innovation exceeds some threshold e.g. when considering costly innovations as soon as individuals can afford to adopt the innovation (e.g. Van der Bulte and Stremersch 2004). It has been shown that both explanations, the social learning explanation described by the Bass model and the heterogeneity explanation, can lead to nearly the same adoption behaviour (Van der Bulte and Stremersch 2004; Kandler and Steele 2010). This again points to the limited power of inferences based on temporal adoption curves.

### 7.2.2 Power-Law Distribution

An important strand of cultural evolution has investigated the utility of models of unbiased social learning or neutral copying, applied mainly in evolutionary archaeology (e.g. Neiman 1995; Shennan and Wilkinson 2001; Bentley and Shennan 2003). The underlying common model is derived through direct analogy to the neutral model of population genetics, where changes in selectively neutral allele frequencies are due simply to the stochastic processes of drift and mutation (e.g. Kimura and Crow 1964). Various empirically calculated diversity statistics of the different archaeological types or variants in an assemblage can be compared with theoretically derived diversity estimators in order to determine whether they are consistent, or not, with the hypothesis of unbiased social learning (Neiman 1995; Kohler

et al. 2004). A further extension of this approach relies on examining the full variant frequency distribution of a dataset, rather than just summary statistics, and comparing it to the right-skewed power-law distribution expected to arise under unbiased social learning. This expectation is again derived from the equivalent model in population genetics, where it was demonstrated that at mutation-drift equilibrium in a finite population the allele frequency distribution is of the power-law form  $P(v) \sim v^{-\alpha}$  (Kimura and Crow 1964), where  $P(v)$  is the proportion of alleles in the population at frequency  $v$ . Such distributions were found across a number of different cultural datasets (e.g. Hahn and Bentley 2003; Herzog et al. 2004; Bentley et al. 2004), and thus the underlying processes in these various cultural systems were deemed to be consistent with an unbiased social learning model. Mesoudi and Lycett (2009) showed that introducing learning biases into the basic neutral model can drive the equilibrium variant frequencies away from a power-law, but relied on a relatively crude testing procedure to identify significant deviation. Clauset and colleagues (2009) provided a very comprehensive set of statistical methods for confirming whether empirical data sets are power-law distributed, demonstrating that the methods adopted, and thus the strength of the conclusions drawn, in the previous literature were necessarily limited.

However, from an information-theoretic viewpoint (e.g. Frank 2009) a power-law distributed dataset is not a firm diagnostic of any particular underlying generative process. The seeming ubiquity of power-law distributions across such a wide range of natural and cultural phenomena, from the magnitude and frequency of earthquakes (Gutenberg and Richter 1944) to word usage frequencies in languages (Zipf 1929), can be explained purely as an inevitable consequence of any large-scale aggregation process that preserves information solely about the geometric mean. In other words, the power-law is a limiting distribution with a large attractor basin, such that invoking a unique mechanism to explain power-law distributed data may be redundant since an infinite number of different mechanisms will generate indistinguishable distributions. Recent work has drawn on this idea by extending the framework outlined by Mesoudi and Lycett (2009), in order to approximately quantify the limits of the power-law attractor basin by testing for when significant deviation from power-law distributed data occurs (Powell et al. *in prep*, see also Clauset et al. 2009). This work begins with an unbiased copying simulation model (e.g. Bentley et al. 2004) and introduces incrementally increasing levels of conformist or anti-conformist social learning. While even very low levels of anti-conformity bias produces clear deviation from a power-law, surprisingly high levels of conformist social learning can occur over many generations without significant deviation from a power-law distribution. This work demonstrates that there is very limited statistical power with

such an approach, and shows the clear limitations of using power-law distributed data as indicative of an underlying unbiased social learning process. Additionally, this power-law approach relies on the assumption that innovation-drift equilibrium is achieved and it is not immediately obvious that such equilibria can be reliably achieved or maintained in many cultural systems, e.g. due to changes in population size or structure.

A closely-related area of work (e.g. Bentley et al. 2007; Acerbi and Bentley 2014) draws on another facet of the neutral model, the rate of turnover within the  $n$  most popular variants in a population, as a further basis to test for divergence from unbiased transmission. Under the unbiased copying assumption this turnover can be well-described in functional form, against which empirical datasets can be examined within a robust model comparison framework (Acerbi and Bentley 2014). This approach appears promising but still relies on the innovation-drift equilibrium assumption and may be difficult to apply it to sparse archaeological data where the total number of observed cultural variants is low (such that a ‘top list’ of size sufficient to accurately calculate turnover cannot be defined).

### 7.2.3 Model Selection Framework

Stressing the existence of a large number of possible social learning strategies, McElreath and co-workers (McElreath et al. 2008) developed a model selection framework based on the Akaike information criterion (AIC) (Akaike 1974; Burnham and Anderson 2002), which allows the quantification of the relative support different learning strategies receive from observed data. To do so they designed an experiment where each individual controls a virtual farm and has to decide which of the two possible crop varieties to plant. Thereby each individual can access information about the pay-off of each crop variety and the frequencies of the different choices in the social group. This set-up produced time-series of decisions and records of which kind of information is used to make this decision. Now in order to use this data to infer the learning strategies employed by the individuals McElreath et al. (2008) developed a probabilistic model which defined the conditional probability of an individual choosing any behavioural option at any point in time given the asocial and social information the individual accessed. Based on these formulations they devised a maximum likelihood framework to fit the considered learning strategies to the observed data. Subsequently the different learning strategies can be compared by evaluating the relative support of the strategies from the data using the AIC. Based on the results of this analysis one can then conclude which learning strategy possesses the highest relative explanatory power among a set of competing and

plausible strategies. McElreath et al. (2008) concluded that a so-called ‘hierarchical compare means/frequency dependent social learning’ strategy receives overwhelming support from the experimental data. This strategy means that individuals compare the mean value of pay-offs of both crop choices and if there is only a small difference they use frequency-dependent social learning to choose which crop to adopt.

This way of approaching the inference problem provides significant advantages over classical hypothesis testing as it can determine which of the strategies, consistent with the data, has the most explanatory power and therefore describes the most likely (however not certain) way learning happened. The ideas developed later in this chapter are in a very similar spirit but we do not restrict the choice of the modelling framework by the availability of an analytical representation of the corresponding likelihood function.

Further, aimed at inferring social learning strategies in animal populations Franz and Nunn (2009) and Hoppitt et al. (2010a) introduced network-based diffusion analysis (NBDA). NBDA explores the existence of social learning in small groups by monitoring whether information flow follows the pattern of association between individuals. The strength of this approach lies in the combination of different kinds of information about the spread process and the social network and an information-theoretic inference framework. NBDA fits simulation models of asocial and social learning to the observed data using maximum-likelihood estimation and infers the underlying learning strategies using model selection based on the AIC (Franz and Nunn 2009). It overcomes a number of the problems associated with the analysis of the adoption curve described in Sect. 7.2.1 but successful applications of this approach require either a controlled experimental set-up or the detailed knowledge of the underlying social network. For a comprehensive review of statistical inference techniques in the animal social learning literature see (Hoppitt and Laland 2013).

### 7.2.4 Formal Modelling

There exists a large body of modelling work focusing on the question of which learning strategies human populations are expected to use especially in spatially and temporally changing environments. It is assumed that cultural variants differ in the benefit they convey to their adopters in a particular environmental situation and population-genetic and game-theoretic analyses revealed the evolutionary stable learning strategies for the assumed environmental fluctuations, i.e. the strategies that would have evolved under natural selection. These equilibria indicated that asocial learning is favoured by natural selection when environmental changes occur on relatively short intervals while social learning is favoured

when the environment is relatively constant (e.g. Aoki et al. 2005; Boyd and Richerson 1985, 1988; Feldman et al. 1996; Rogers 1988). Other approaches such as the simulation approach of the ‘social learning strategies tournament’ (Rendell et al. 2010, 2011) or the  $n$ -variant competition framework (Kandler and Laland 2013) suggested that social learning could be the dominant learning strategy even in rapidly changing environmental conditions. This difference is partly caused by the fact that in these more recent studies more than two cultural variants, offering different levels of benefit in a particular environment, are considered. The potential accumulation of social information then allows social learners to adjust their adopted variant after an environmental change by switching to an available, more adaptive variant.

Further, social learning can occur in a large number of different ways (e.g. Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985; Laland 2004), which raises the question of how social information should be used to adapt efficiently to changing environmental conditions. In this context special attention has been devoted to conformist social learning, in which common variants are disproportionately adopted (Boyd and Richerson 1985). Henrich and Boyd (1998) studied the conditions under which conformity can evolve and found that selection favours conformist learning as long as the environment does not change too rapidly and the evolution of social learning is more strongly influenced by environmental heterogeneity than the evolution of conformity. Contrary to those results Nakahashi (2007), Wakano and Aoki (2007), Kendal et al. (2009) and Kandler and Laland (2013) demonstrated a negative relationship between environmental stability and the reliance on conformist social learning: the reliance on conformity tends to be larger in frequently changing environments. Wakano and Aoki (2007) pointed out that the Henrich and Boyd model results in the same conclusion when the time of evolution is sufficiently long. Additionally the models mentioned above differ in their assumptions, e.g. Henrich and Boyd (1998) and Wakano and Aoki (2007) used mixed learning strategies in a spatial and non-spatial setting, respectively whereas Nakahashi (2007), Kendal et al. (2009) and Kandler and Laland (2013) assumed pure strategies where individuals use either social or asocial learning. Eriksson et al. (2007) suggested that allowing for more than two cultural variant (and additionally that individuals have complete knowledge about all present variants) will disadvantage the evolution of conformist social learning. Further, Efferson et al. (2008) noted that the evolutionary advantage of the conformist strategy depends on the accuracy of asocial learning. McElreath et al. (2011) pointed to the importance of spatial heterogeneity in evolutionary modelling and claimed that non-spatial frameworks may underestimate the role of conformity. Similarly, Nakahashi et al. (2012) argued that the focus on (unrealistic) situations characterized by: (i) only two cultural variants present, (ii) only temporally

varying environments and (iii) error-free social learning has obscured conditions favouring the evolution of conformity.

In summary, much theoretical work has been devoted to the analysis of evolutionary stable strategies in different environmental conditions. This approach revealed a large number of important insights into the evolutionary process: It allows for an elegant characterisation of the long-term outcomes of evolution through natural selection as well as of the associated evolutionary trajectories and for the derivation of analytical conditions for different evolutionary trajectories. In other words, it produces predictions of which learning strategies are expected to be present in the population given the environmental circumstances. However, in many real-world situations learning strategies cannot be observed directly and therefore the model predictions reviewed above are difficult to verify with the available data. Further, the analysis of stable evolutionary equilibria implicitly assumes that sufficient time has passed so that the equilibrium could be reached and that individual fitness (with respect to the considered cultural trait) is the quantity maximised by evolution (Maynard Smith 1978). But how do we know that the considered cultural system is at equilibrium? Are those stable evolutionary equilibria really an appropriate description of which learning strategies human populations use?

### 7.3 Generative Inference

In order to address the inverse problem of inferring underlying learning strategies that could have given rise to observed data, there exist two major types of modelling approaches: discriminative and generative models. Discriminative inference proceeds by fitting a, usually limited, number of model parameters conditional on a single instance of empirical data, generally assuming some correlational relationship between these observations and the implicit parameters, which may have no explicit or causal bearing on the true underlying mechanism. This kind of approach includes many standard and familiar statistical modelling procedures, such as regression or clustering techniques, in which parameter values can be estimated or structurally different models be compared. On the other hand, the generative approach proceeds by building a fully specified probabilistic model, in which the hypothesised causal mechanisms are explicitly defined. This model is used to repeatedly simulate pseudo-datasets under known parameter values such that their expected distribution can then be statistically compared with the observed data. This comparison allows certain hypothesised mechanisms to be rejected as inconsistent with the empirical data, or the estimation of the causal model parameters that provide the best fit to the data.

The idea of generative inference in socio-cultural evolution really first emerged with the introduction of ‘Sug-



arscape', one of the earliest agent-based models designed to investigate social processes (Epstein and Axtell 1996). By repeatedly simulating a population of autonomous agents moving and interacting on a heterogeneous resource landscape the model demonstrated that large-scale social phenomena or patterns, analogous to those observed in the real world, could emerge from the aggregation of micro-level interactions. 'Sugarscape' and subsequent work (see also Epstein 2007) demonstrated the viability of this inference approach, by investigating the effects of hypothesised and explicitly defined individual-level causal mechanisms on population-level outcomes, that could then be compared with observed data in order to test models and estimate parameter values. While one of the major advantages of this approach is that expected outcomes of much more realistic, and potentially non-equilibrium, systems can easily be generated, the principle limitation has been the lack of a robust statistical methodology capable of comparing the outcomes from these complex models to empirical data. Much of the work discussed in Sect. 7.2.2 and some in Sect. 7.2.1 can be considered as taking this generative approach, but the inferences drawn are limited by the lack of power in the statistical methodology employed. However, a relatively recent major statistical advance in the form of approximate Bayesian computation (driven mainly by statistical geneticists, e.g. Beaumont et al. 2002) has started to transform the field of population genetics, by allowing inference on increasingly detailed generative genetic models of human evolutionary history using observed genomic data. Its recent successes should serve as a model for the field of cultural evolution, and we are strongly advocating this approach here.

### 7.3.1 Application in Population Genetics

A significant area of population genetics has recently emerged that relies on the generative inference approach outlined above. This work is usually in the form of explicit simulation modelling, in which simulated genetic data sets are generated under increasingly realistic genetic (McVean and Cardin 2005; Marjoram and Wall 2006) and/or demographic (e.g. Currat et al. 2004) models, and then statistically compared to observed data in order to infer population history (Veeramah et al. 2012; Eriksson et al. 2012) or past selection pressures (Itan et al. 2009; Wilde et al. 2014).

But from its origins in the early twentieth century (Fisher 1930; Wright 1931) up until fairly recently, population genetics had been a largely *prospective* discipline (Ewens 2004), concerned with predicting how allele frequencies in a population will change over time. These mainly analytical models were used to identify theoretical equilibria

between the processes of mutation, drift and selection and the expected levels of genetic variation thus maintained in a population, as well as predicting the long-term fate of new alleles. However, an alternative way of viewing evolution is based on the idea of genealogies, or lineages, where the ancestral relationships of a contemporary sample of individuals are considered *retrospectively*. One major advantage of this approach is that the ancestry of non-sampled individuals can be disregarded, as, by definition, it has no bearing on the evolution of the sampled individuals. While in genetics these ideas date back at least to the 1940s (e.g. Malécot 1948), it was only the development of coalescent theory in the 1980s (Kingman 1982) that allowed the ancestral process to be robustly embedded within a population. The coalescent is a statistical model that describes the stochastic process by which lineages reach common ancestors, or 'coalesce', backwards in time within a population of defined size or structure, generating a dated genealogy rooted by the most recent common ancestor of the whole sample. Many classical population genetics results were re-derived, often more elegantly, using this retrospective approach, but the real strength of the coalescent is that it allows extremely efficient simulation of genealogies under a well-defined population history without having to assume any sort of equilibrium. In conjunction with a mutation model, these genealogies can then be used to rapidly generate simulated genetic data to match each observed sample. A number of extensions to the basic coalescent model have allowed the incorporation of evolutionary processes such as recombination (Hudson 1983; Marjoram and Wall 2006), as well as complex or geographically realistic demographic histories (e.g. Currat et al. 2004; Eriksson et al. 2012).

The coalescent approach is used (almost exclusively) for inference based on neutral genetic loci, in order to avoid the serious confounding effects of natural selection. Disentangling the effects of demography and selection can be very difficult, especially in human populations, given that many genetic adaptations in humans are thought to have been driven by cultural processes, e.g. adaptations due to dietary changes (see Laland et al. 2010). It is difficult to extend the coalescent to such complex cases, so instead genetic data can be simulated by modelling the whole population forward in time, and explicitly defining selection pressures that may vary in time and/or space or apply just to certain cultural groups (e.g. Itan et al. 2009; Wilde et al. 2014). This *forward simulation* approach can be far more computationally demanding, but still the underlying principle is the same: genetic data is generated under an explicitly defined evolutionary model which does not need to assume equilibrium. Either way, these simulated genetic samples can then be compared to observed samples, in order to estimate the parameters of the evolutionary causal mechanisms, such

as selection strength, changes in population size or migration, or test between different models of population history.

This simulation, or generative inference, approach has only become fully viable since the development of an extremely powerful statistical methodology known as approximate Bayesian computation (ABC) (Beaumont et al. 2002). While it is generally preferable to base statistical inferences on a likelihood function, in most applied population genetic models their explicit specification and calculation can be challenging, if not impossible. For those (often simplistic) models in which the likelihood can be specified, approaches such as Markov chain Monte Carlo (MCMC) can be implemented in order to evaluate the likelihood. Such methods are, in general, extremely computationally demanding and it can be difficult to ensure that the chain converges on the target distribution efficiently. In contrast, ABC allows statistically robust inferences to be made on complex models without the consideration of explicit likelihood functions. The general idea is to approximate the joint posterior distribution of the model parameters  $P(\theta|D)$  directly, given observed data  $D$ . This is done by repeatedly simulating data  $D^*$  under a generative model with parameter values drawn from their prior distributions  $P(\theta)$ , and retaining just those parameter sets that generate data sufficiently ‘close’ to the real data. These retained parameter sets form a sample from the distribution  $P(\theta|d(D, D^*) \leq \varepsilon)$ , where  $d(\dots)$  is the distance between the observed and simulated data and  $\varepsilon$  is tolerance level that determines the level of approximation. As  $\varepsilon \rightarrow 0$  the retained sample asymptotically approaches the true posterior distribution  $P(\theta|D)$ . In practice, the data are usually reduced in dimension by using a set of summary statistics  $S$ , such that we are approximating the posterior  $P(\theta|S)$  with  $P(\theta|d(S, S^*) \leq \varepsilon)$ . Parameter estimation is then performed via the joint or marginal posteriors for each parameter, such that modal values and credible intervals can be obtained. Many subsequent developments of this basic algorithm have allowed successful applications of ABC to a wide variety of inference problems both in population genetics and elsewhere (e.g. Toni et al. 2009, and see Sect. 7.3.2.2).

Recent debate in population genetics has reinforced the necessity of a rigorous model-based approach to inference (Nielsen and Beaumont 2009; Beaumont et al. 2010; Templeton 2009). This allows explicit parameterized models to be created, tested and fitted to data, in a way that other approaches, such as phylogeography, cannot. While a number of other non-generative model-based methods are producing exciting findings (e.g. Green et al. 2010; Hellenthal et al. 2014), an increasing number of researchers are utilising the generative approach, often coupled with ABC, to answer diverse questions about human evolutionary history using both standard population genetic (Veeramah

et al. 2012; Eriksson et al. 2012) as well as gene-culture coevolutionary (Itan et al. 2009; Rasteiro et al. 2012) models.

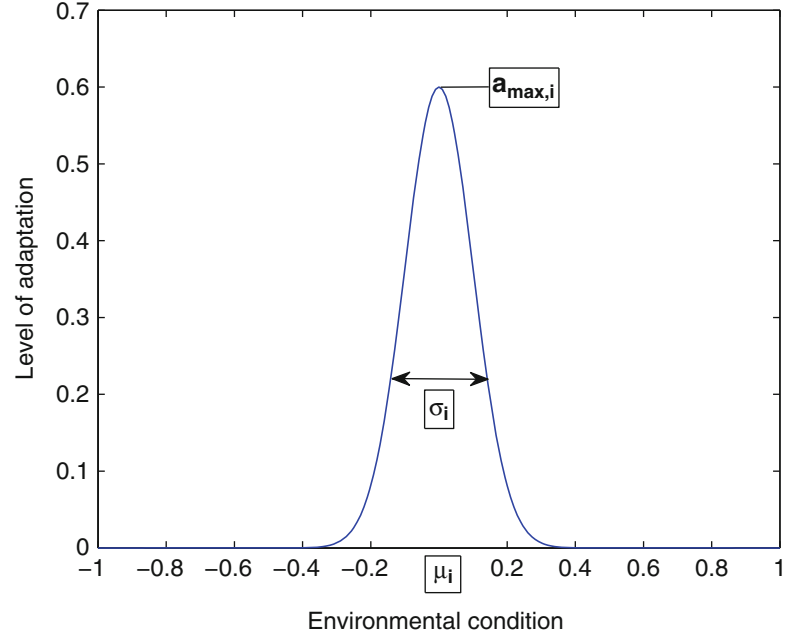
### 7.3.2 Application to Cultural Evolution

In this section we turn back to our initial question of which learning strategies human populations use especially in changing environments. Based on the framework developed by Kandler and Laland (2013) we demonstrate in the following how a cultural evolutionary generative inference procedure can be constructed and how it can be used to explore which learning hypotheses are consistent with the empirical evidence. Thereby we assume that time series data detailing the usage or occurrence frequencies of different cultural variants are available. Similarly to the genetic applications the inference procedure consists of two steps. The first step comprises the development of a non-equilibrium generative model capturing the main cultural and demographic dynamics of the considered system and describing the frequency changes of different cultural variants present in a population under the assumed learning hypothesis. The focus on non-equilibrium situations is motivated by observations of e.g. drastic environmental changes or temporal variations in the total size of the population of cultural variants. The model is designed to establish a causal relationship between the various asocial and social learning strategies and temporal frequency changes of the present cultural variants. Consequently initializing the model with the frequencies observed at the beginning of the time series produces patterns of frequency change expected under a specific learning hypothesis at any future point in time. The second step uses ABC to derive conclusions about which mixtures of learning strategies are consistent with the observable frequency data and which are not. The aim of this framework, then, is to allow researchers to ‘reverse engineer’ which learning strategies are likely to have been used in current or past populations, given knowledge of how frequencies have changed over time, independent of any optimality or equilibrium assumptions. It should be noted that the generative model we introduce below is simply an example to demonstrate the utility of this approach, and that a number of the models reviewed in Sect. 7.2 could feasibly be adapted to use within this framework. The crucial point is that the generative model is appropriately chosen and adjusted to the considered application.

#### 7.3.2.1 Generative Model

In Kandler and Laland (2013) we considered  $k$  different variants of a cultural trait that differ in the benefit they

**Fig. 7.1** Example of an adaptation function



convey to their adopters in a particular environment. Using a competition framework we modelled the frequency changes of those  $k$  variants of a cultural trait in the face of asocial and social learning in form of direct biased and conformist social learning. We assumed that a population is experiencing changing environmental conditions, expressed by the function  $e(t)$  with  $e(t) \in [-1, 1]$ . Those changing conditions affect the adaptation levels of the different cultural variants and we characterised each variant  $i$  by its adaptation function  $a_i(e(t))$ . This function indicates the benefit the variant conveys to its adopters in condition  $e(t)$  and is determined by three parameters: the environmental condition  $\mu_i$  to which the variant is best adapted to, the maximum level of adaptation  $a_{\max,i}$  and the width  $\sigma_i$  (see Fig. 7.1 for an example).

Now asocial and social learning strategies cause the cultural variants to change in frequency and we described those changes using a differential equation framework of the form

$$\begin{aligned} \frac{\partial u_i}{\partial t} &= -\nu u_i(t) + \xi \text{ (asocial learning)} \\ &\quad + (1 - \xi) \text{ (social learning)}, \quad i = 1, \dots, k \\ \frac{dK}{dt} &= (\lambda - \nu) \times K(t) (1 - K(t)) \end{aligned} \quad (7.1)$$

The variable  $u_i$  describes the frequency of variant  $i$  at time  $t$  in the population, or in other words the fraction of the population that has adopted variant  $i$ . Further,  $K$  denotes the total population size,  $\lambda$  and  $\nu$  the birth and death rates, respectively and  $\xi$  the reliance of the population on asocial learning. It follows from the equation describing the tem-

poral dynamic of the population size in model (7.1) that  $K(t) \leq 1$ ,  $\forall t$  and further it holds  $\sum_{i=1}^k u_i(t) \leq K(t)$ . Consequently, the temporal change in frequency of variant  $i$  (denoted by  $du_i/dt$ ) is determined by

- the loss of adopters of variant  $i$  (described by  $\nu u_i$ ) and
- the cumulative influences of social and asocial learning (described by the terms  $\xi$  (asocial learning) +  $(1 - \xi)$  (social learning)).

Social learning is based on social cues (and therefore can only lead to learning of variants which are already present in the considered location), however the nature of the cues can be ambiguous. We considered two different social learning strategies: direct biased social learning and conformist social learning. Direct biased social learning is based on adaptation information (meaning the adaptation levels  $a_i$  of the cultural variants in the current environment  $e(t)$ ) and defined as the adoption of variants found to be efficacious by individual assessment (Boyd and Richerson 1985). This is incorporated in model (7.1) by assuming the following dynamic: The higher the adaptation level of a cultural variant the stronger is its transmission supported by direct biased social learning. Contrary, conformist social learning is based solely on frequency information and leads to a disproportional adoption of variants whose frequencies are above a commonness threshold (Boyd and Richerson 1985). It holds: the higher the variant's frequency the stronger it is supported by the conformist bias. The reliance on adaptation information and frequency information is modulated by the parameter  $b$  where  $b = 0$  indicates no conformist tendencies. Asocial learning on the other hand is based on environmental cues, specifically on the inference about the benefit of specific



variants in observed environmental conditions. Naturally, this inference is error-prone but can lead to the introduction of new variants into a specific location. We note that the considered cultural system consist of at most  $k$  different variants; variants can possess a frequency of zero at times and be reintroduced into the system by the process of asocial learning. In order to explore the temporal dynamic of the learning process the birth and death rates,  $\lambda$  and  $\nu$ , are kept constant. A detailed mathematical description of the considered model can be found in the Appendix. We note that the suggested framework can be easily extended to a spatially explicit model by allowing the adopters of the different cultural variants to move in space. These spatial movements can be described by a simple diffusion process and the appendix provides the description of a spatially explicit model.

Summarizing, system (7.1) provides a framework that generates temporal frequency patterns of different cultural variants possessing different levels of adaptation to the experienced environment conditioned on the assumed mixture of learning strategies, i.e. it assumes that a fraction  $\xi$  of the population engages in asocial learning and consequently a fraction  $(1 - \xi)$  engages in social learning whereby the strength of conformist learning is determined by the parameter  $b$ . The adaptation level  $a_i$  of the present cultural variants is determined by experienced environmental conditions  $e(t)$  and the specific adaptation functions as given in Fig. 7.1. Importantly, the environment  $e(t)$  can vary in a deterministic or random manner (for more details see Kandler and Laland 2013). Now a systematic variation of the model parameters  $(\xi, b)$  in system (7.1) produces insights into the interplay between learning strategies and patterns of frequency changes (see (Kandler and Laland 2013) for a detailed analysis). In particular when initializing the developed generative model with the frequency data observed at the first time point we obtain expectations about possible frequency patterns at later time points conditioned on the assumed learning hypothesis  $(\xi, b)$  and adaptation levels  $a_i$  of the variants. In the following we aim to establish statistically which of the possible mixtures of learning strategies are consistent with the observed frequency data.

### 7.3.2.2 Statistical Inference

As already mentioned above, in order to explore which mixtures of learning strategies are consistent with the observed data we would ideally determine the likelihood function of system (7.1) indicating the likelihood of the assumed parameter values  $\theta = [a_i, \xi, b]$  given the observed data  $D$ . In this framework the data consists of frequency information about different cultural variants at different points in time. Based on the likelihood function we could derive the region of parameter space which is likely to have produced the

observed frequency patterns and therefore conclude which mixtures of the learning strategies could adequately explain the data. However, in our case the likelihood function cannot be determined easily and so we use ABC in the form of a Sequential Monte Carlo (SMC ABC) algorithm (Toni et al. 2009; Beaumont 2010; Del Moral et al. 2012) to approximate the posterior distribution  $P(\theta|D)$ . The key idea of the SMC ABC algorithm is to decompose this problem into a series of simpler sub-problems as efficient sampling from approximate posterior distribution  $P(\theta|d(D, D^*) \leq \varepsilon)$  can prove to be difficult. The algorithm starts by choosing a sequence of tolerance levels  $\{\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n\}$  with  $\varepsilon_0 < \varepsilon_1 < \dots < \varepsilon_n \geq 0$ . In the first step a sample is drawn from  $P(\theta|d(D, D^*) \leq \varepsilon_0)$  with  $\varepsilon_0$  large and subsequently from an increasingly constrained sequence of distributions  $P(\theta|d(D, D^*) \leq \varepsilon_i)$ ,  $i = 1, 2, \dots, n$ . In other words the tolerance level is sequentially decreased until the desired level  $\varepsilon = \varepsilon_n$  is reached. In this context, Del Moral et al. (2012) introduced a procedure which adaptively determines the best tolerance level  $\varepsilon_i$  in each step.

The SMC ABC algorithm produces distributions of the model parameters that indicate the ranges of the parameter space which are able to produce frequency data within a given tolerance margin  $\varepsilon$  of the observed data. Consequently, the mixture of learning strategies that are consistent with the data can be deduced. As suggested in Kandler and Laland (2013) the widths of the marginal posterior distributions will allow an estimation of the amount of information contained in the data on each parameter. If the ranges are narrow then only a small region of the parameter space is consistent with the data and therefore we conclude that a large number of mixtures of learning strategies are *not* able to produce the observed frequency change. In this way the suggested approach can help reduce the possible range of strategies that could have produced the observed frequency data. On the other hand, if the posterior distributions of the model parameters are wide a large region of the parameter space is consistent with the data. This might be indicative of either an inadequate description of the cultural system by the generative model or the fact that the observed frequency data do not carry a strong signature of the underlying learning strategies: many different mixtures of learning strategies can result in very similar frequency changes and therefore not much information about the underlying learning processes can be gained from analysing frequency data. Even though this appears as a negative result it is crucial to understand when frequency data is *not* informative about the underlying evolutionary processes, as in this situation researchers have to base inference on other lines of evidence.

### 7.3.2.3 Example

In order to demonstrate the applicability of the suggested framework we consider the following theoretical example. We use frequency data generated by the previously described generative model (7.1) in the time interval  $[0,100]$  and sample  $n$  data points as shown in Fig. 7.2. In more detail, we assume three cultural variants to be present initially (with initial fractions of adopters  $u_1(0) = 0.2$ ,  $u_2(0) = 0.1$  and  $u_3(0) = 0.07$ ) and the environment to stay constant until time  $t = 50$ . Then a shock occurs causing the environment to change, the present variants to decrease in frequency and two new variants to be introduced. The model parameters are chosen as follows

$\xi = 0.1$ ,  $b = 0$ ,  $a_1 = 0.3$ ,  $a_2 = 0.45$ ,  $a_3 = 0.4$  for the time interval  $[0,50]$  and  
 $\xi = 0.1$ ,  $b = 0$ ,  $a_1 = 0.35$ ,  $a_2 = 0.1$ ,  $a_3 = 0.07$ ,  
 $a_4 = 0.6$ ,  $a_5 = 0.45$  for the time interval  $[50,100]$ .

Using this approach we have full knowledge of the system (i.e. the adaptation level of the variants, learning processes and environmental fluctuations) that generated the data, and this enables us to explore how well our proposed inference framework recovers the true parameters from disturbed data (see coloured squares in Fig. 7.2). The data is disturbed by adding Gaussian noise distributed as  $\mathcal{N}(0, \sigma^2)$  with  $\sigma^2 = 0.03^2$ .

Now we apply the suggested inference framework to the noisy frequency data and aim to estimate the adaptation levels of variants present in the time intervals  $[0,50]$  and  $[50,100]$ , respectively, as well as the fraction of asocial learners  $\xi$  and the strength of conformist learning strategy  $b$ . Figure 7.3 shows the resulting posterior distribution for  $\xi$  and  $b$  for the interval  $[0,50]$  (top row) and the interval  $[50,100]$  (bottom row) for  $n = 11$  data points. These distributions describe the range of the parameter space of  $\xi$  and  $b$  where the squared sum of the differences between the theoretical and observed frequencies is smaller than  $\varepsilon$ . Thereby  $\varepsilon$  is chosen to be the squared sum of the differences of disturbed and undisturbed data as shown in Fig. 7.2 (Toni et al. 2009). It is obvious that in both cases the posterior distributions are tightly clustered around the true value  $\bar{\xi} = 0.1$  and  $\bar{b} = 0$  and therefore the underlying learning processes can be recovered relatively accurately from the noisy data. For sake of shortness we omitted the posterior distributions for the adaptation level  $a_i$ , however note that those are also tightly clustered around the true values. Figure 7.3c, f show the joint distributions of the parameter  $\xi$  and  $b$  in the different time intervals. The light areas indicate areas with high probability whereas the darker areas indicate less likely parameter regions. The joint distributions provide valuable information about the interactions of the investigated processes, in our case asocial learning and conformist social learning. In other words, they provide an indication of which parameter con-

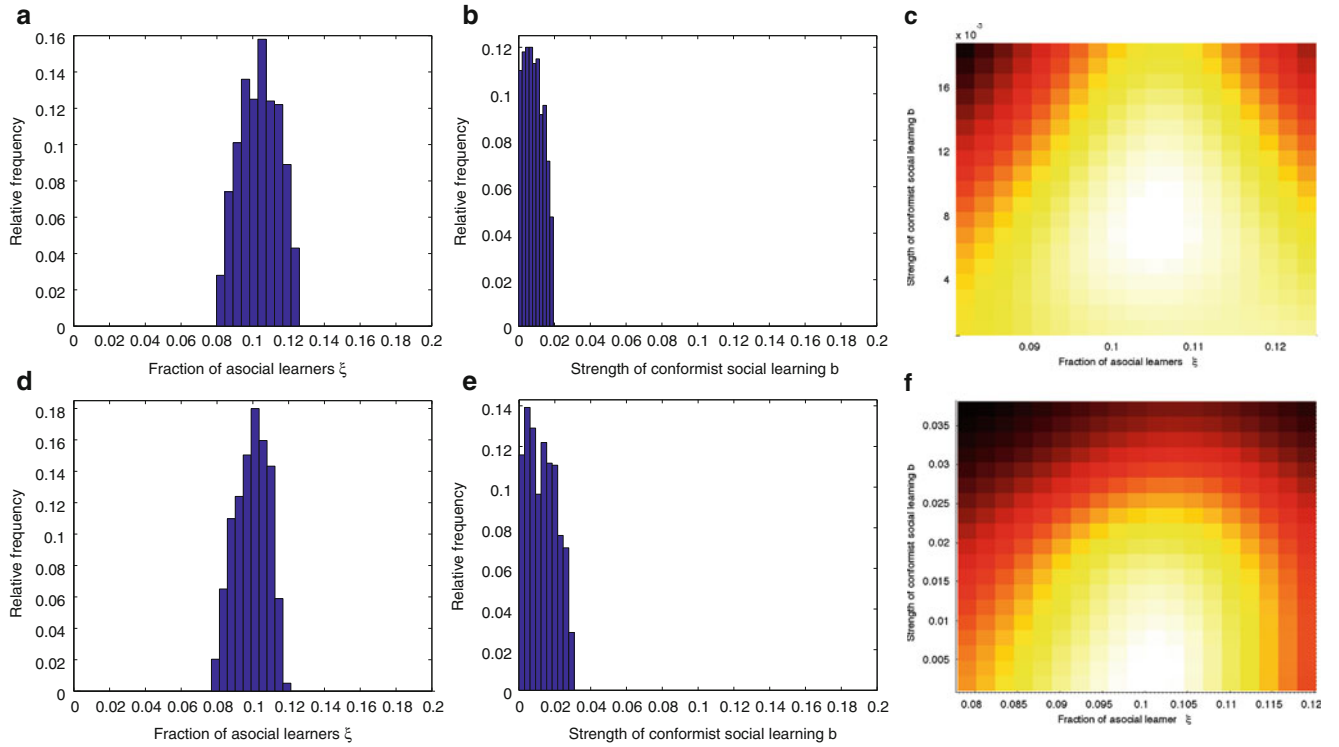
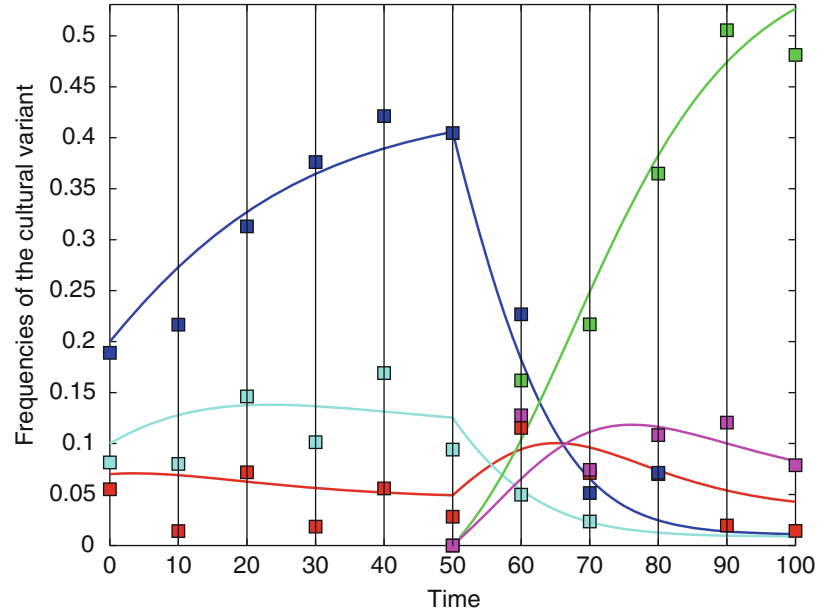
stellation give rise to similar population-level frequency data. It is obvious that large deviations from the true value in both parameters do not produce frequency patterns consistent with the observed data (see the triangular shapes of the light areas in Fig. 7.3c, f). We conclude that the developed inference framework is able to recover the true model parameter on the base of  $n = 11$  data points.

This set-up also allows us to explore the question of how many data points are needed to obtain reliable inference results. Figure 7.4 shows the posterior distributions for  $\xi$  and  $b$  for the time interval  $[0,50]$  if only  $n = 3$  data points at  $t = 0;25;50$  (top row) and  $n = 2$  data points at  $t = 0;50$  (bottom row) are available. Not surprisingly the ranges of consistent parameter values are broadened and therefore more mixtures of learning processes are consistent with the sparse data. Especially if only two data points are available we cannot make conclusive inferences about the role of conformist social learning. The range  $[0,0.08]$  of the parameter space for  $b$  is consistent with the data which means that we cannot exclude the presence of a weak or medium-strength conformist bias in the population. Nevertheless based on only two data points, we can firstly reject the hypothesis of the existence of a strong conformist tendency and secondly establish that the fraction of asocial learner is between 4 and 18 %. Further, the joint posterior distribution (Fig. 7.4f) shows that a stronger tendency for conformist social learning is accompanied by a larger fraction of asocial learning in the population (see shape of the light region). In other words, increasing the strength of conformist social learning  $b$  and the fraction of asocial learning  $\xi$  in the populations simultaneously can lead to very similar frequencies of the three present cultural variants at  $t = 50$ .

## 7.4 Conclusion

Model-based inference approaches have already started to transform the field of population genetics, by allowing inference on increasingly detailed generative genetic models of human evolutionary history using observed genomic data. Acknowledging that fundamental questions in cultural evolution and population genetics are of similar nature, namely which underlying evolutionary mechanisms could have given rise to observed frequency distributions, we explored in this chapter the applicability of this inference technique to cultural data. The key to success of such a framework is the development of a generative model capturing the main demographic and cultural dynamics of the considered cultural system. This means that properties like population structure, changes in population size or patterns of environmental change are input to the model and need to be appropriately chosen. Here we considered a competition framework modelling frequency changes of  $k$  different variants of a cultural

**Fig. 7.2** Time course of the frequencies of the different variants

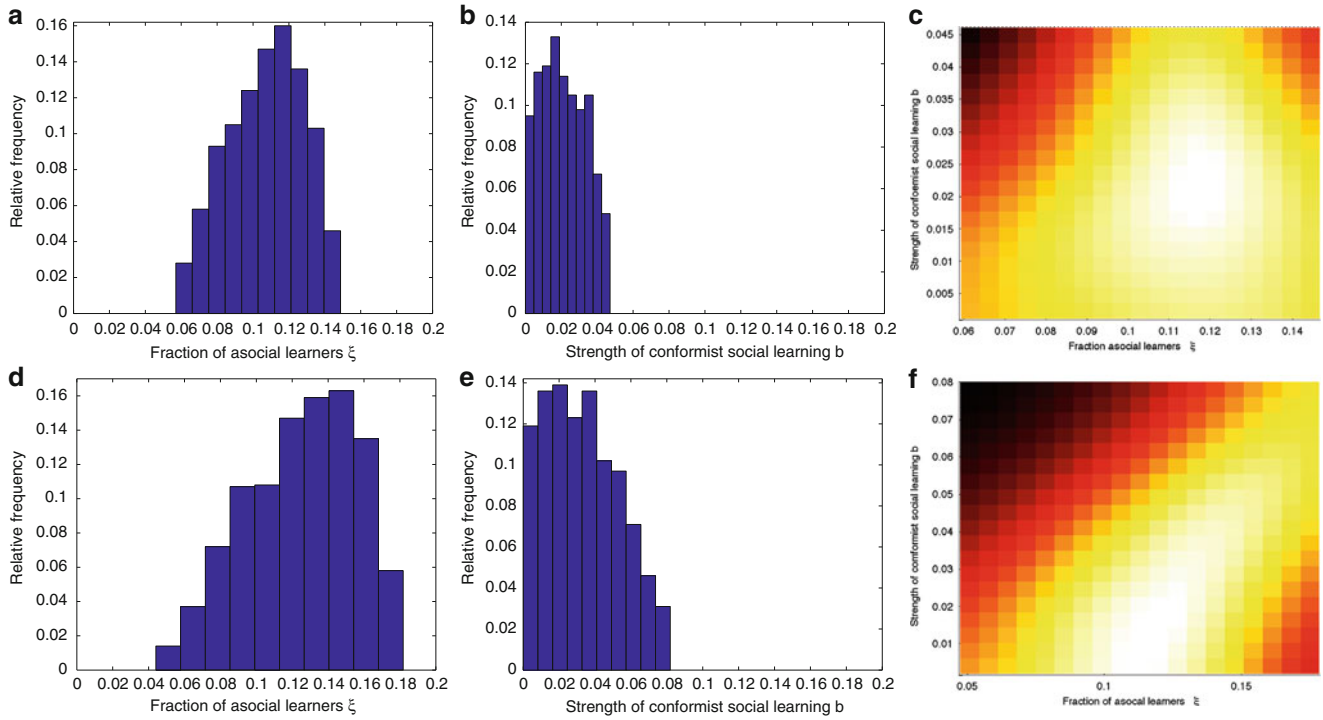


**Fig. 7.3** Posterior distributions for  $\xi$ ,  $b$  and the corresponding joint posterior distribution for the interval  $[0, 50]$  (a–c), posterior distributions for  $\xi$ ,  $b$  and the corresponding joint posterior distribution for the interval

$[50, 100]$  (d–f). In all cases  $n = 11$  data points are sampled and the true parameter values are  $\xi = 0.1$  and  $b = 0$

trait in the face of asocial and social learning in the form of direct biased and conformist learning in a spatially and temporally variable environment. We note that this chosen framework is only an example of a generative model; other approaches such as simulation or network-based frameworks can be used in exactly the same way (see Crema et al. 2014

for an application of a simulation based inference framework applied to archaeological data). The crucial point is that the generative model firstly produces data to which observed data can be compared to and secondly establishes a causal relationship between the processes under investigation (in our case the employed learning process) and the generated



**Fig. 7.4** Posterior distributions for  $\xi$ ,  $b$  and the corresponding joint posterior distribution for the interval  $[0,50]$  when  $n = 3$  (a–c), posterior distributions for  $\xi$ ,  $b$  and the corresponding joint posterior distribution

for the interval  $[0,50]$  when  $n = 2$  (d–f). The true parameter values are  $\bar{\xi} = 0.1$  and  $\bar{b} = 0$

data so that statistical inference procedures such as ABC methods can be applied. In this way the suggested procedure does not depend on any equilibrium assumption and infers conclusions from a comparison of theoretical and observed patterns of frequency change.

While conventional hypothesis testing is able to establish whether the observed data are consistent with a single hypothesis but does not allow for conclusions about the range of hypotheses that are consistent, the ABC methodology produces posterior distributions for the model parameters indicating the parameter ranges that are consistent with the observed data. The widths of those posterior distributions provide estimates of how uniquely the observed population-level data can be generated from a specifically defined mixture of learning strategies. In other words, this framework should help to address the long-standing question of how much information about the underlying evolutionary processes is really contained in observed population-level data. While narrow posterior distributions for the model parameter describing the strength of the different learning processes indicate that the observed patterns of frequency change can only be produced by a small number of hypothesis wide distributions do not allow for informative conclusions about underlying learning processes based on the considered data and the generative model. We expect the developed infer-

ence framework to be valuable in helping to narrow down the range of possible hypotheses that could have produced observed data and therefore to be instructive especially in the face of sparse data.

Additionally, careful application of an ABC model selection framework can also potentially allow robust comparison between structurally different hypotheses. Under the Bayesian paradigm competing models can be evaluated on the basis of Bayes factors. While implicitly penalising over-parameterized complex models these factors serve as a summary of the evidence provided by the data in favour of a specific cultural hypothesis out of the possibilities considered. Early applications of this idea in connection with ABC (e.g. Pritchard et al. 1999) used the notion that Bayes factors can be empirically approximated by calculating the posterior probabilities of the competing models,  $i$ ,  $P(\theta^i | D)$ . In practice, this is achieved by assigning each competing hypothesis equal prior probability and then, given a fixed tolerance or error margin  $\varepsilon$ , approximating model posterior probabilities by their relative acceptance rates in the rejection algorithm (Toni and Stumpf 2010). In our example application we would distinguish between two models: our original model (7.1) accounting for asocial learning and social learning in the form of direct biased and conformist learning and a less complex model which does not allow for

a conformist tendency in the population (This is achieved by setting  $b = 0$  in model (7.1)). As expected we find that the less complex model has a higher Bayes factor as the original data were generated under the assumption of no conformity in the population.

In the application presented in Sect. 7.3.2.3 we compared theoretical and observed frequency data  $D$  directly in the SMC ABC algorithm, using the squared distance between the data points. But in most real applications using the data directly might not be possible, and then the ABC framework relies on reducing the full data  $D$  to a set of summary statistics  $S$  such that the approximated posterior distribution  $P(\theta | S \leq \varepsilon)$  is adequately representative of  $P(\theta | D \leq \varepsilon)$ . In statistical terms this requires  $S$  to be ‘sufficient’ for  $D$ , such that no information is lost in the dimension reduction. In practice this is almost impossible to guarantee, so choosing  $S$  can be difficult. A number of different heuristic schemes to guide the choice of  $S$  have been proposed, such that  $S$  ‘approaches sufficiency’ (e.g. Joyce and Marjoram 2008) or choice is otherwise optimized (e.g. Nunes and Balding 2010) or automated (e.g. Fearnhead and Prangle 2012). In addition, the model selection approach described above can become problematic when the Bayes factors are approximated on the basis of summary statistics  $S$  rather than the full data  $D$  (e.g. Robert et al. 2011), as even if  $S$  is sufficient for each model it may not be sufficient to compare between competing models. This is currently a topic of intense research in both population genetics and statistics in general, and regular improvements allow these limitation to be somewhat mitigated in careful analyses. Notwithstanding these limitations, the ABC family of techniques has in just a short time permitted a revolutionary change in inferential power using increasingly complex and realistic models, especially those in population genetics. However, we stress that the accuracy of the obtained inference depends crucially on the appropriate description of the considered cultural or genetic system by the generative model. Naturally this points to the importance of the understanding of the nature and complexity of the learning strategies employed by human populations.

In order to infer which learning strategies human populations really use we need to base inference on available empirical evidence. Large individual-level datasets on social learning are relatively rare, outside of experimental conditions. However, two recent examples have drawn on detailed historical records of the board game Go (Beheim et al. 2014) or contemporary ethnographic data from Fiji (Henrich and Broesch 2011) to directly infer the relative importance of different learning-biases across and how this may impact population-level cultural dynamics. While both studies clearly demonstrate the advantages of such high-

resolution data, in most pre-modern contexts (including the palaeolithic period) we are necessarily limited to aggregate population-level data. The framework we have outlined here successfully demonstrates that we can often still use such sparse data to make inferences on the relative weights of various underlying individual-level learning processes. Importantly, the framework also indicates when the available population-level data do not carry a strong signature about underlying learning processes and therefore when the analysis of this data will not provide meaningful information. The ABC inference framework offers an alternative way of studying cultural phenomena, independent of any equilibrium or optimality assumptions and is highly appropriate for application to typical archaeological datasets (see also Crema et al. 2014). In particular, this approach provides a useful bridge between the important on-going theoretical work in cultural evolution and the continually growing empirical data on human evolutionary history.

**Acknowledgement** We thank Kenichi Aoki and Jeremy Kendal for their constructive comments which helped improving this chapter.

## A.1 Appendix

In this chapter we assume that asocial and social learning strategies cause the cultural variants to change in frequency and describe those changes using a reaction–diffusion framework of the form

$$\begin{aligned} \frac{\partial u_i}{\partial t}(t, x) &= d \Delta u_i(t, x) - \nu u_i(t, x) + \xi \text{ (asocial learning)} \\ &\quad + (1 - \xi) \text{ (social learning)}, \quad i = 1, \dots, k \\ \frac{\partial K}{\partial t}(t, x) &= d \Delta K(t, x) + (\lambda - \nu) \\ &\quad \times K(t, x) (1 - K(t, x)). \end{aligned} \quad (7.2)$$

Thereby the variable  $u_i$  describes the frequency of variant  $i$  at time  $t$  in the population, or in other words the fraction of the population that has adopted variant  $i$ . The variable  $K$  denotes the population size at location  $x$  and in the following we assume the population size to be the same for all locations  $x$ . It follows from the second equation in model (7.2) that  $K(t, x) \leq 1$ ,  $\forall t$  and further, it holds  $\sum_{i=1}^k u_i(t, x) \leq K(t, x)$ . The diffusion coefficient  $d$  describes the scale of spatial interactions,  $\lambda$  and  $\nu$  the birth and death rates, respectively and  $\xi$  the reliance of the population on asocial learning. For sake of simplicity we stated the non-spatial version in the main text. All dynamics described below hold in a similar way for this model.



In more detail, learning in various forms can increase or decrease the frequency of variant  $i$ . Asocial learning is based on the judgement about the benefit of specific variants in observed environmental conditions and consequently has two error sources: misjudgement of the current environmental condition and misjudgement of the adaptation levels of the different variants. Despite the conceptual differences both error sources lead to the same outcome in the modelling framework: a variant  $i$  is chosen for which holds  $\mu_i \neq e$ . Therefore the inaccuracy of asocial learning is modelled by assuming that asocial learning is based on  $\bar{e} = e + \omega$  with  $\omega \sim \mathcal{N}(0, \sigma_\omega^2)$ . However, besides being error-prone asocial learning can lead to the introduction of new variants and its dynamic is modelled by (for sake of shortness we write  $\bar{e} = \bar{e}(t, x)$ )

$$\begin{aligned} & P_i(a_i(\bar{e})) \left( K(t, x) - \sum_{j=1}^k u_j \right) \\ & + \sum_{j=1, j \neq i}^k (P_{ji}(a_i(\bar{e}), a_j(\bar{e})) u_j(t, x) \\ & - P_{ij}(a_i(\bar{e}), a_j(\bar{e})) u_i(t, x)) \end{aligned} \quad (7.3)$$

The parameter  $P_i$  describes the rate at which the fraction of the population which has not yet adopted a variants (described by the difference between the population size  $K(t, x)$  at time  $t$  and the sum of the fractions of the population which have adopted one of the  $k$  variants,  $K(t, x) - \sum_{j=1}^k u_j(t, x)$ ) learns variant  $i$  asocially.  $P_i$  depends on the adaption level  $a_i(\bar{e})$  meaning that asocial learning is not completely random: the higher the adaptation level in the estimated environment  $\bar{e}$  the higher is the adoption rate. Further, we allowed for the switching of variants which describes the process that individuals who already have adopted a cultural variant can switch to adopting a different variant. The coefficient  $P_{ij}$  models the rate at which the fraction of the population which has adopted variant  $i$  switches to variant  $j$  due to the evaluation of environmental cues. Again it holds the larger the difference  $a_j(\bar{e}) - a_i(\bar{e})$  between the estimated adaption levels the higher is the switching rate. Contrary to asocial learning, social learning is based on social cues and therefore can only lead to learning of variants which are already present in the considered location. In the considered framework we only considered two social learning strategies: direct biased social learning and conformist social learning. Direct biased social learning is modelled by (for sake of shortness we write  $e = e(t, x)$ )

$$\begin{aligned} & r_i(a_i(e)) u_i(t, x) \left( 1 - \frac{u_i(t, x)}{K(t, x) - \sum_{j=1, j \neq i}^k u_j(t, x)} \right) \\ & + \sum_{j=1, j \neq i}^k (c_{ji}(a_i(e), a_j(e)) \\ & - c_{ij}(a_i(e), a_j(e))) u_i(t, x) u_j(t, x). \end{aligned} \quad (7.4)$$

Similarly to the dynamic of asocial learning the first term

$$r_i(a_i(e)) u_i(t, x) \left( 1 - \frac{u_i(t, x)}{K(t, x) - \sum_{j=1, j \neq i}^k u_j(t, x)} \right)$$

models the adoption of variant  $i$  by the population which has not adopted any variants yet. However contrary to asocial learning, this term is frequency-dependent. It is a logistic growth process with adoption rate (or intrinsic rate of increase)  $r_i$  and broadly speaking describes cultural reproduction. Per definition, the population size  $K(t, x)$  at location  $x$  is the upper limit of the total fraction of adopters in the population at this location (given by  $\sum_{j=1}^k u_j(t, x)$ ), regardless of the adopted variant. Consequently, the upper limit for the fraction of the population that has adopted variants  $i$  is given by  $K(t, x) - \sum_{j=1}^k u_j(t, x)$  (i.e. we assume that our cultural variants compete for a common pool of adopters). The adoption rate  $r_i$  is assumed to be proportional to the adaptation level  $a_i$  in the currently experienced environmental condition  $e$ . It holds: The higher the adaptation level the higher is the adoption rate. The second term

$$\begin{aligned} & \sum_{j=1, j \neq i}^k (c_{ji}(a_i(e), a_j(e)) \\ & - c_{ij}(a_i(e), a_j(e))) u_i(t, x) u_j(t, x) \end{aligned}$$

describes the switching dynamic between the fractions of the population which has already adopted a variant. Again we assumed that individuals who have already adopted a variant have the chance to switch to another variant and therefore the different cultural variants compete with each other for use. These interactions between the variants are described by the terms  $c_{ij}(a_i(e), a_j(e)) u_i(t, x) u_j(t, x)$  which model the switch process from variant  $i$  to variant  $j$ . The strength of this process is determined by the rate  $c_{ij}$  and it holds: The higher the difference  $a_j(e) - a_i(e)$  of the adaptation levels of both variants the higher is the switching rate. In order to

include conformist social learning we allowed these model parameters to be frequency-dependent. We assumed

$$\begin{aligned}\tilde{r}_i &= (1 - b) r_i(a_i(e)) + b(u_i(t, x) - c_b K(t, x)) \text{ and} \\ \tilde{c}_{ij} &= [(1 - b) c_{ij}(a_i(e), a_j(e)) + b(u_j(t, x) \\ &\quad - c_b K(t, x))]^+\end{aligned}$$

where  $b$  controls the reliance on adaptation information and frequency information, respectively. For  $b = 0$  we obtain direct biased learning while  $b > 0$  supports variants with a frequency higher than the commonness threshold  $c_b K(t, x)$ . In this case the difference  $(u_i(t, x) - c_b K(t, x))$  is positive and the adoption rate  $\tilde{r}_i$  is increased. Contrary if  $(u_i(t, x) - c_b K(t, x))$  is negative (and therefore variant  $i$  has a relatively small frequency) the adoption rate  $\tilde{r}_i$  is decreased. A similar dynamic applies to the switching rate  $\tilde{c}_{ij}$ . If the frequency of variant  $j$  (the target of the switch process) exceeds the commonness threshold  $c_b K(t, x)$  then the rate  $\tilde{c}_{ij}$  with which variant  $i$  is substituted by variant  $j$  is increased. The symbol  $[\cdot]^+$  denotes the positive part of any real number (e.g.  $[3.4]^+ = 3.4$  but  $[-3.4]^+ = 0$ ) ensures that there is no reversal of the switch direction.

We note that when considering a single cultural variant the dynamic of asocial learning (7.3) results in  $r$ -shaped adoption curve while the dynamic of asocial learning (7.4) results in a  $S$ -shaped curve whereby the existence of a conformist tendency ( $b > 0$ ) produces long tails at the beginning and an accelerated adoption behaviour when the commonness threshold is exceeded. System (7.2) can be solved using the Finite-Element method and we obtain the time course of the frequencies  $u_i$  of each cultural variant that are expected under the assumed learning hypothesis and environmental change.

## References

- Acerbi A, Bentley RA (2014) Biases in cultural transmission shape the turnover of popular traits. *Evol Hum Behav* 35:228–236
- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Aoki K, Wakano JY, Feldman MW (2005) The emergence of social learning in a temporally changing environment: a theoretical model. *Curr Anthropol* 46:334–340
- Bass FM (1969) A new product growth model for consumer durables. *Manag Sci* 15:215–227
- Baum WM, Richerson PJ, Efferson CM, Paciotti BM (2004) Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. *Evol Hum Behav* 25:305–326
- Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol Evol Syst* 41:379–406
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035
- Beaumont MA, Nielsen R, Robert C, Hey J, Gaggiotti O, Knowles L, Estoup A, Panchal M, Corander J, Hickerson M, Sisson SA, Fagundes N, Chikhi L, Beerli P, Vitalis R, Cornuet J-M, Huelsenbeck J, Foll M, Yang Z, Rousset F, Balding D, Excoffier L (2010) In defence of model-based inference in phylogeography. *Mol Ecol* 19:436–446
- Beheim BA, Thigpen C, McElreath R (2014) Strategic social learning and the population dynamics of human behavior: the game of Go. *Evol Hum Behav* 35:351–357. doi:10.1016/j.evolhumbehav.2014.04.001
- Bentley RA, Shennan SJ (2003) Cultural transmission and stochastic network growth. *Am Antiq* 68:459–485
- Bentley RA, Hahn MW, Shennan SJ (2004) Random drift and culture change. *Proc R Soc B* 271:1443–1450
- Bentley RA, Lipo CP, Herzog HA, Hahn MW (2007) Regular rates of popular culture change reflect random copying. *Evol Hum Behav* 28:151–158
- Boyd R, Richerson PJ (1985) Culture and the evolutionary process. The University of Chicago Press, Chicago
- Boyd R, Richerson PJ (1988) An evolutionary model of social learning: the effects of spatial and temporal variation. In: Zentall T, Galef BG Jr (eds) Social learning. Erlbaum, Hillsdale, pp 29–48
- Burnham K, Anderson D (2002) Model selection and multi-model inference: a practical information-theoretic approach. Springer, Berlin
- Caldwell CA, Millen AE (2008) Experimental models for testing hypotheses about cumulative cultural evolution. *Evol Hum Behav* 29:165–171
- Cavalli-Sforza L, Feldman MW (1981) Cultural transmission and evolution: a quantitative approach. Princeton University Press, Princeton
- Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *Soc Ind Appl Math Rev* 51(4):661–703
- Coultas JC (2004) When in Rome ... an evolutionary perspective on conformity. *Group Process Intergroup Relations* 7(4):317–331
- Crema ER, Edinborough K, Kerig T, Shennan SJ (2014) An approximate Bayesian computation approach for inferring patterns of cultural evolutionary change. *J Archaeol Sci* 50:160–170
- Curat M, Ray N, Excoffier L (2004) Splatche: a program to simulate genetic diversity taking into account environmental heterogeneity. *Mol Ecol Notes* 4(1):139–142
- Del Moral P, Doucet A, Jasra A (2012) An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat Comput* 22(5):1009–1020
- Efferson C, Lalive R, Richerson PJ, McElreath R, Lubell M (2008) Conformists and mavericks: the empirics of frequency-dependent cultural transmission. *Evol Hum Behav* 29:56–64
- Epstein JM (2007) Generative social science: studies in agent-based computational modeling. Princeton University Press, Princeton
- Epstein JM, Axtell RL (1996) Growing artificial societies: social science from the bottom up. MIT and Brookings Institution, Washington, DC
- Eriksson K, Enquist M, Ghirlanda S (2007) Critical points in current theory of conformist social learning. *J Evol Psychol* 5(1–4): 67–87
- Eriksson A, Betti L, Friend AD, Lycett SJ, Singarayer JS, von Cramon-Taubadel N, Valdes PJ, Balloux F, Manica A (2012) Late Pleistocene climate change and the global expansion of anatomically modern humans. *Proc Natl Acad Sci U S A* 109:16089–16094
- Ewens WJ (2004) Mathematical population genetics, 2nd edn. Springer, New York
- Fearnhead P, Prangle D (2012) Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *J R Stat Soc Ser B* 74:419–474

- Feldman MW, Aoki K, Kumm J (1996) Individual versus social learning: evolutionary analysis in a fluctuating environment. *Anthropol Sci* 104:209–232
- Fisher RA (1930) *The genetical theory of natural selection*. Clarendon, Oxford
- Frank SA (2009) The common patterns of nature. *J Evol Biol* 22:1563–1585
- Franz M, Nunn CL (2009) Network-based diffusion analysis: a new method for detecting social learning. *Proc R Soc B* 276(1663):1829–1836
- Giraldeau L-A, Valone TJ, Templeton JJ (2002) Potential disadvantages of using socially acquired information. *Philos Trans R Soc Lond B* 357:1559–1566. doi:10.1098/rstb.2002.1065
- Green RE, Krause J, Briggs AW et al (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722
- Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. *Bull Seismol Soc Am* 34:185–188
- Hahn MW, Bentley RA (2003) Drift as a mechanism for cultural change: an example from baby names. *Proc R Soc B* 270:120–123
- Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S (2014) A genetic atlas of human admixture history. *Science* 343(6172):747–751
- Henrich J (2001) Cultural transmission and the diffusion of innovations: adoption dynamic indicate that biased cultural transmission is the predominant force in behavioural change. *Am Anthropol* 103(4):992–1013
- Henrich J, Boyd R (1998) The evolution of conformist transmission and the emergence of between-group differences. *Evol Hum Behav* 19:215–241
- Henrich J, Broesch J (2011) On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. *Philos Trans R Soc B* 366:1139–1148
- Herzog HA, Bentley RA, Hahn MW (2004) Random drift and large shifts in popularity of dog breeds. *Proc R Soc B* 271:1443–1450
- Heyes CM (1994) Social learning in animals: categories and mechanisms. *Biol Rev* 69:207–231
- Hoppitt WJE, Laland KN (2013) *Social learning: an introduction to mechanisms, methods, and models*. Princeton University Press, Princeton
- Hoppitt WJE, Boogert NJ, Laland KN (2010a) Detecting social transmission in networks. *J Theor Biol* 263(4):544–555
- Hoppitt WJE, Kandler A, Kendal JR, Laland KN (2010b) The effect of task structure on diffusion dynamics: implications for diffusion curve and network-based analyses. *Learn Behav* 38(3):243–251
- Hudson RR (1983) Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol* 23:183–201
- Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG (2009) The origins of lactase persistence in Europe. *PLoS Comput Biol* 5:e1000491
- Joyce P, Marjoram P (2008) Approximately sufficient statistics and Bayesian computation. *Stat Appl Genet Mol Biol* 7:26
- Kandler A, Laland KN (2013) Tradeoffs between the strength of conformity and number of conformists in variable environments. *J Theor Biol* 332:191–202
- Kandler A, Steele J (2010) Social learning, economic inequality and innovation diffusion. In: O'Brien MJ, Shennan S (eds) *Innovation in cultural systems*. The MIT Press, Cambridge, MA, pp 193–214
- Kendal JR, Kendal RL, Laland KN (2007) Quantifying and modelling social learning processes in monkey populations. *Int J Psychol Psychol Ther* 7(2):123–138
- Kendal JR, Giraldeau L-A, Laland KN (2009) The evolution of social learning rules: payoff-biased and frequency-dependent biased transmission. *J Theor Biol* 260:210–219
- Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49:725–738
- Kingman JFC (1982) The coalescent. *Stochastic Process Appl* 13:235–248
- Kirby S, Cornish H, Smith K (2008) Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc Natl Acad Sci U S A* 105(31):10681–10686
- Kohler TA, Van Buskirk S, Ruscavage-Barz S (2004) Vessels and villages: evidence for conformist transmission in early village aggregations on the Pajarito Plateau, New Mexico. *J Anthropol Archaeol* 23:100–118
- Laland KN (2004) Social learning strategies. *Learn Behav* 32(1):4–14
- Laland KN, Richerson PJ, Boyd R (1996) Developing a theory of animal social learning. In: Heyes CM, Galef BG Jr (eds) *Social learning in animals: the roots of culture*. Academic Press, London, pp 129–154
- Laland KN, Odling-Smee FJ, Myles S (2010) How culture shaped the human genome: bringing genetics and the human sciences together. *Nat Rev Genet* 11:137–148
- Mahajan V, Muller E, Bass FM (1995) Diffusion of new products: empirical generalizations and managerial uses. *Mark Sci* 14(3):79–88
- Malécot G (1948) *The mathematics of heredity* (trans: Yermanos DM 1969). WH Freeman, San Francisco
- Marjoram P, Wall JD (2006) Fast “coalescent” simulation. *BMC Genet* 7:16
- Maynard Smith J (1978) Optimization theory in evolution. *Annu Rev Ecol Syst* 9:31–56
- McElreath R, Bell AV, Efferson C, Lubell M, Richerson PJ, Waring T (2008) Beyond existence and aiming outside the laboratory: estimating frequency-dependent and pay-off-biased social learning strategies. *Philos Trans R Soc B* 363:3515–3528
- McElreath R, Fasolo B, Wallin A (2011) The evolutionary rationality of social learning. In: Hertwig R, Hoffrage U (eds) *Simple heuristics in a social world*. Oxford University Press, New York
- McVean GAT, Cardin NJ (2005) Approximating the coalescent with recombination. *Philos Trans R Soc B* 360:1387–1393
- Mesoudi A, Lycett SJ (2009) Random copying, frequency-dependent copying and culture change. *Evol Hum Behav* 30(1):41–48
- Mesoudi A, O'Brien MJ (2008) The cultural transmission of great basin projectile-point technology I: an experimental simulation. *Am Antiq* 73(1):3–28
- Morgan TJH, Rendell L, Ehn M, Hoppitt WJE, Laland KN (2012) The evolutionary basis of human social learning. *Proc R Soc B* 279:653–662
- Nakahashi W (2007) The evolution of conformist transmission in social learning when the environment changes periodically. *Theor Popul Biol* 72:52–66
- Nakahashi W, Wakano JY, Henrich J (2012) Adaptive social learning strategies in temporally and spatially varying environments. *Hum Nat* 23:386–418
- Neiman FD (1995) Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in Illinois woodland ceramic assemblages. *Am Antiq* 60:7–36
- Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Mol Ecol* 18(6):1034–1047
- Nunes MA, Balding DJ (2010) On optimal selection of summary statistics for approximate Bayesian computation. *Stat Appl Genet Mol Biol* 9:34
- Powell A, Shennan SJ, Thomas MG (in prep) The power of power-laws in cultural evolution
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol* 16(12):1791–1798
- Rasteiro R, Bouttier P-A, Sousa V, Chikhi L (2012) Investigating sex-biased migration during the Neolithic transition in Europe. *Proc R Soc B* 279:2409–2416

- Reader SM (2004) Distinguishing social and asocial learning using diffusion dynamics. *Anim Learn Behav* 32(1):90–104
- Rendell L, Boyd R, Cowden D, Enquist M, Eriksson K, Feldman MW, Fogarty L, Ghirlanda S, Lillicrap T, Laland KN (2010) Why copy others? Insights from the social learning strategies tournament. *Science* 328(5975):208–213
- Rendell L, Fogarty L, Hoppitt WJE, Morgan TJH, Webster MM, Laland KN (2011) Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends Cogn Sci* 15(2):68–76
- Robert CP, Corneut J-M, Marin J-M, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proc Natl Acad Sci U S A* 108:15112–15117
- Rogers AR (1988) Does biology constrain culture? *Am Anthropol* 90:819–831
- Rogers EM (2003) *Diffusion of innovations*, 5th edn. Free Press, New York
- Shennan SJ, Wilkinson JR (2001) Ceramic style change and neutral evolution: a case study from Neolithic Europe. *Am Antiq* 66:577–594
- Templeton AR (2009) Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographical analysis vs. approximate Bayesian computation. *Mol Ecol* 18:319–331
- Toni T, Stumpf MPF (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26(1):104–110
- Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPF (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* 6:187–202
- Van der Bulte C, Stremersch S (2004) Social contagion and income heterogeneity in new product diffusion: a meta-analytic test. *Mark Sci* 23(4):530–544
- Veeramah KR, Wegmann D, Woerner A, Mendez F, Watkins J, Destro-Bisol G, Soodyall H, Louie L, Hammer MF (2012) An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Mol Biol Evol* 29:617–630
- Wakano JY, Aoki K (2007) Do social learning and conformist bias coevolve? Henrich and Boyd revisited. *Theor Popul Biol* 72: 504–512
- Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterlaender M, Hollfelder N, Potekhina ID, Schier W, Thomas MG, Burger J (2014) Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000y. *Proc Natl Acad Sci U S A* 111(13):4832–4837
- Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159
- Zipf GK (1929) Relative frequency as a determinant of phonetic change. *Harv Stud Class Philol* 15:1–95