University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

2018

Null Models For Cultural And Social Evolution

Mitchell Newberry
University of Pennsylvania, mgnew@sas.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Applied Mathematics Commons, Biostatistics Commons, and the Evolution Commons

Null Models For Cultural And Social Evolution

Abstract

Analogies between biological and cultural evolution date back to Darwin, yet the analogies have remained loose. Neutral evolution, known to be important in biology, has been

proposed as a null model for cultural change, but has not developed into tests for selection on cultural features. Using inference in timeseries of alternative word forms and

grammatical constructions, I demonstrate a cultural analog of natural selection on a background of netural evolution. Social evolution, on the other hand, implies selection in a

social environment and therefore cannot be described with a neutral model. I propose a

model of pure frequency-dependent selection as a generic null model for social evolution,

and use the model to illustrate diverse effects of social selection. I derive a non-linear

form of frequency-dependent selection from a mechanistic model of mate choice and show

unintuitive consequences for evolutionary dynamics. I infer complex forms of frequency

dependent selection—including positive and negative frequency-dependent selection at different frequencies—from data regarding the copying of baby names, the fashions of dog

breeds, and the use of rare languages, and discuss the implications for cultural diversity.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biology

First Advisor
Joshua B. Plotkin

Second Advisor
Paul D. Sniegowski

Keywords
exchangeability, fashion, frequency-dependent selection, language, neutrality, standardization

Subject Categories
Applied Mathematics | Biostatistics | Evolution

NULL MODELS FOR CULTURAL AND SOCIAL EVOLUTION

Mitchell G. Newberry

A DISSERTATION

in

Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

_____

Joshua B. Plotkin, Professor of Biology

Graduate Group Chairperson

_____

Michael Lampson, Associate Professor of Biology

Dissertation Committee

Paul Sniegowski (chair), Professor of Biology and Stephen A. Levin Family Dean of the College of Arts and Sciences

Erol Akçay, Assistant Professor of Biology

Timothy Linksvayer, Associate Professor of Biology

Arjun Raj, Assistant Professor of Bioengineering

NULL MODELS FOR CULTURAL AND SOCIAL EVOLUTION

ACKNOWLEDGEMENT

ABSTRACT

NULL MODELS FOR CULTURAL AND SOCIAL EVOLUTION

Mitchell G. Newberry

Joshua B. Plotkin

Analogies between biological and cultural evolution date back to Darwin, yet the analogies have remained loose. Neutral evolution, known to be important in biology, has been proposed as a null model for cultural change, but has not developed into tests for selection on cultural features. Using inference in timeseries of alternative word forms and grammatical constructions, I demonstrate a cultural analog of natural selection on a background of netural evolution. Social evolution, on the other hand, implies selection in a social environment and therefore cannot be described with a neutral model. I propose a model of pure frequency-dependent selection as a generic null model for social evolution, and use the model to illustrate diverse effects of social selection. I derive a non-linear form of frequency-dependent selection from a mechanistic model of mate choice and show unintuitive consequences for evolutionary dynamics. I infer complex forms of frequency dependent selection—including positive and negative frequency-dependent selection at different frequencies—from data regarding the copying of baby names, the fashions of dog breeds, and the use of rare languages, and discuss the implications for cultural diversity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

PREFACE

Perhaps I can save you time, dear reader, by providing context. This thesis is a presentation of three papers deemed satisfactory to represent the contribution of my doctoral studies. Two of these papers (here called Chapter 1 and Chapter 2) are already published (Newberry et al., 2016, 2017). These are best apprehended as published in their respective venues but are included here *pro forma*. The third chapter is a manuscript under preparation to be submitted to some academic journal or another. If by the time you read this, that work is already published, you will likely have an easier time reading it in that venue, or indeed on bioRxiv or arxiv.org. What is uniquely expressed in this presentation is the motivation and the intellectual thread that connects and underlies these works, including some of its loose ends and sketches of its implications and my future work. In the interest of clarity, then, all of the synthesis of the joint body of work is confined to the introduction and conclusion, and the chapters are presented in chronological order of publication, edited only superficially from their published or yet-to-be published form.

The thesis began as a study of how populations of people and organisms come to cohere on certain features. That is, if the endless copying inherent to the evolutionary process constantly produces variation, how do we nonetheless agree on anything, particularly in cases where the content of the agreement seems inconsequential, such as whether to measure in inches or centimeters, or whether to wait 33 or 34 years to flower. The scope of the thesis since expanded to include the complementary case of a preference for disagreement and anything in between, as well as a study of how to make sense of the analogy between genetic and cultural variants in order to study these processes generally in biology as well as in human culture.

Each of the papers presented consist substantially of my original writing and original research. The exceptions are original research on which I have collaborated at such a deep level as to know each line of code and the origin of each piece of input data, to have partici-

pated and weighed in on each decision of any consequence, and to easily rearrange, replicate and reconfigure at will. These are intensive collaborations with my advisor Joshua Plotkin crafting every lilt and turn of word, paragraph, and structure of our paper in *Nature*, with Chris Ahern on all the software machinery underlying the same, and with David McCandlish on the introduction in the style of *The American Naturalist* of Chapter 1, which was never submitted to *Am Nat*.

INTRODUCTION

The idea embodied here is simple: to explore a certain class of evolutionary models because those are the easy ones. We often model evolution essentially as Lotka or Volterra modeled growth of populations. Such a model represents a population or an ecosystem as population counts of a number of different types that change over time, where the growth rate of each type depends on its abundance and the abundance of other types. These models are notoriously complex, from either a mathematical or a practical perspective. Chaotic dynamics can result from including even a handful of types. Even though the Lotka-Volterra model excels at representing the interactions of two types, say a predator and its prey, generalizing them to an ecological food web of n types requires n^2 parameters, which are usually impossible to measure or infer in practice. Mathematically, we can sketch out the full space of possible Lotka-Volterra models, which represent a wide swath of possible ecosystems and evolutionary scenarios. What we see is that among all possibilities, only vanishingly few of them could be comprehensible enough to be at all predictable or fit to any real data. Mapping out these vanishing few, then, should be of central concern.

I study the models where the growth rate of each type depends only on its own abundance and the total abundance, and each in the same way. This is a fairly straightforward restriction to specify, and it reduces the full space of possibilities to a smaller subspace. The simple idea of the thesis, then, is that the restriction introduces a symmetry, which in turn guarantees that the behavior of any such model is predictable and can be fit to real data. The restriction allows us to elaborate models of populations not just in a deterministic or large-population limit as Lotka and Volterra did, but in full stochastic generality—making account of the randomness inherent to population change. Furthermore, the ecological and evolutionary scenarios that this subspace of models represent are interesting, because the subspace captures competition between demographically-equivalent types, which includes the neutral model of molecular evolution and the neutral theory of biodiversity as well as models of frequency-dependent selection. This thesis explores the models themselves and

their applications to evolutionary scenarios, about which much more could be said.

Neutrality is a kind of demographic equivalence that has been studied in evolution and ecology. Neutral models in evolutionary theory—where alternative traits have no consequence for an organism's fitness—were developed in the early to mid 20th century (Wright, 1929; Kimura, 1983) and demonstrated a non-adaptive dimension to evolutionary change. More recently, a neutral theory of ecology (Hubbell, 2001) has been put forth to explain biodiversity patterns. Neutral models are useful as null models, to explore first whether the idiosyncrasies of particular types or traits are relevant at all to evolutionary or ecological dynamics. If an observed evolutionary change seems plausible under a neutral model, then there might be no need to resort to a more complex explanation involving natural selection or other processes favoring the growth of one or another type. Likewise if the abundance of different species in an assemblage of species is adequately explained by random colonization and proliferation, then a detailed understanding of their relationships and ecological niches may be unnecessary for predicting biodiversity patterns. These null expectations, then, have been useful topics of study in their own right. This work extends neutral null models to the domain of cultural transmission, and proposes a new null model of frequency-dependent selection for social evolution.

Although cultural and social evolution might sound similar, I use these terms to refer to different and nearly independent aspects of the evolutionary process. Cultural transmission refers to the process of individuals aquiring traits from other individuals through experience such as by teaching, learning, immitating, or hearing and repeating, rather than through genetic inheritance. Words (Creanza et al., 2015), bird songs, bird flight paths (Sasaki and Biro, 2017) and tool use in primates (Tomasello et al., 1987) are examples of identifiable behavioral traits that are transmitted culturally. Where these traits are transmitted iteratively from one individual to another, the process parallels Darwinian genetic evolution and can be called cultural evolution (Cavalli-Sforza and Feldman, 1981; Mesoudi, 2011). Social evolution on the other hand often refers to situations in which the fitness

2

of a trait is affected by the "social" interaction—interaction between individuals withing the same population that depends on their respective traits. Example traits include alarm calling behavior (Dunford, 1977), flowering time (Janzen, 1976) and the production of enzymes for extracellular metabolism in microbes (Greig and Travisano, 2004). A trait need not be culturally-transmitted in order to be subject to selection from the social environment, as in microbes (Kerr et al., 2002), nor necessarily subject to social selection to be culturally-transmitted, as in stone tool construction (Lycett, 2008, 2009). Cultural and genetic transmission can entwine to create social selective pressures as well, such as in co-evolution of dairy practice—a culturally-transmitted trait—and lactase enzyme expression, which is inherited genetically (Itan et al., 2009). Likewise cultural transmission and social selection go hand in hand: we often unwittingly conflate social and cultural evolution simply because the vast majority of culturally-transmitted traits germaine to our everyday human experience have meaning only in a social context, and are therefore under social selection in one or another sense.

In Chapter 2, I extend the neutral model from population genetics to the context of culturally-transmitted traits by using the neutral model as a null model to distinguish "selective" forces from neutral change. The idea that cultural transmission should entail a system of copying that bears some analogy to Darwinian evolution by natural selection is at least as old as Darwin's work (Darwin, 1888), and was popular at the time (Aronoff, 2017). In particular August Schleicher saw parallels between species and language evolution. Schleicher, a linguist, reconstructed the a family tree of Indo-European languages (Schleicher, 1861), and is famous for Schleicher's Fabel, the first plausible myth written in a plausible phylogeny-based reconstruction of the extinct progenitor of all Indo-European languages. A contemporary of Darwin's, Schleicher observed lines of descent in language that he saw as analogous to linneages of species. Schleicher explicitly described the analogy under the provocative title *Darwinism Tested by Language Science* (Schleicher, 1869), which was quickly translated from the German. A review of the translated work by a philologist appears in the first volume of *Nature*. Müller (1870) writes in the review,

"A struggle for life is constantly going on among the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their own inherent virtue."

which Darwin quotes in *The Descent of Man* (Darwin, 1888, p.91).

How far one can take the analogy between cultural and genetic transmission has been the topic of contemporary research and debate (Lewens, 2015). Cavalli-Sforza and Feldman (1981) produced a detailed mathematical framework for studying evolution by cultural transmission, in particular to describe gene-culture coevolution as exemplified in the relationship between lactase persistence genes and dairy cultures, or genes and language attributes (Creanza et al., 2015). Yet wide adoption of Darwinism as an explanation for observed changes or patterns has been slow in the social sciences. A principle difficulty is in determining what precisely the analogy is, particularly as regards the "genes" of culture. It is clear that some information that determines traits such as behaviors, words or attributes of technological artifacts, is recursively transmitted, but what constitutes the "cultural replicator", the embodiment of that information that would be analogous to DNA in genetic transmission, is often unclear or context-specific. All the same, evolutionary thinking in biology predated the discovery of DNA, and so it is reasonable to expect that just as evolution shed light on biology before the exact genetic mechanisms were understood, we can hope for evolution to explain aspects of cultural change even when detailed mechanistic knowledge of the encoding of cultural information or its mechanisms of transmission are unavailable.

The many types of cultural information and the many ways in which it can be transmitted ramify questions about individuality and levels of selection which were vexing even in the biological contexts that first prompted their study. Whether genes copy themselves using the organism as a vehicle, whether individuals reproduce depending on the aggregate fitness of their genes, or indeed whether groups of individuals produce individuals according to some group fitness have long puzzled evolutionary biologists (West et al., 2007; Dawkins, 1976). Furthermore, situations arise in biology where the notion of individuality underlying

Darwinian fitness breaks down and causes paradoxes in evolutionary thinking (Buss, 2014). For example the case of free-living amoebae coming together to form multicellular structures at reproduction presents the mystery of how evolution could stably maintain the traits required to become multicellular when most individual cells do not benefit from the added reproductive capacity.

Culture forces reckoning with even more complex notions of individuality. For example, do word forms themselves have a cultural fitness, or do they confer biological fitness to speakers, or confer cultural "fitness" to speakers' idiolects or languages? Mufwene (2001) ascribes to languages the properties of viruses or parasites that infect hosts. For most purposes a host either speaks a given language or does not, much the same way a human is infected with a parasite or not, and the meaningful evolutionary dynamics of parasites have as much to do with the process of transmission between hosts as with the survival and reproduction of individual parasite cells or virus particles. The nuanced evolutionary dynamics of this patch colonization might then be quite complicated.

A gambit helps to sidestep these questions in order to extend the neutral model to cultural transmission. Much as a phenotypic gambit (Grafen, 1984) permits simple modeling of the evolutionary outcomes of social strategies, the present gambit permits us to ignore questions about the cultural replicator: we simply model observed traits as if they were the individuals under study. For example, I take the occurrence of a word in a corpus (a collection of texts) to represent the free-living cultural replicator itself. Even though the occurrence of a word in the corpus may represent many physical printed words in many copies of a book on many shelves, or memories in the brains of many readers, these other pools are taken as irrelevant to they dynamics of cultural change in much the same way that the course of an epidemic can be modeled without knowing the exact number of virus particles within each infected person's body or even whether the disease is viral or bacterial. A full justification of this gambit would have to answer to all possible transmission mechanisms that might relay cultural information in all possible contexts, and is therefore likely impossible. We

instead rely on strong justifications along mathematical, empirical and practical lines.

The mathematical justification centers on the diffusion approximation, a theme that pervades this thesis. The idea is that when there are many individuals and many possible interactions between individuals that are all or mostly uncoordinated, the frequencies of traits can only change slightly within a short time interval. If we further assume that the possible changes depend only on the current or very recent states of the population, then nearly any process can be approximated by a stochastic process called the diffusion process. The diffusion process requires specification of only how the mean and variance of a change depend on the current state of the system. The fact that only the mean and variance matter is related to the Central Limit Theorem by the assumption that there are many individuals and interactions, and thus their aggregate effect is approximately normal. This normal approximation allows us to ignore many of the details of what the real individuals and their interactions are, and classify these processes only according to their aggregate effect. Indeed, models in population genetics differ in terms of how a population and the mechanisms of survival and reproduction are represented, yet they share the same diffusion limit: that is, they have approximately the same behavior when the relevant populations and timescales are large (Crow et al., 1970; Kimura, 1964). For example, the Moran process represents instantaneous birth and death whereas the Wright-Fisher process represents birth and death over discrete generations, yet the behavior of trait frequencies in both models is approximate to a shared diffusion limit, and the effective population size in the diffusion limit allows interconversion of results between the models. Many complicated models thus may be summarized in terms of how the instantaneous mean and variance of trait frequency depend on the current trait frequencies, to good approximation. This fact allows the study of the system-level behavior of transmission and evolution by studying factors that affect this instantaneous mean and variance, to which the detailed mechanisms are mathematically irrelevant.

An empirical and practical justification is that the aggregate behavior of many cultural

variants seem superficially consistent with a diffusion process (Bentley et al., 2004; Hahn and Bentley, 2003; Mesoudi and Lycett, 2009; Bentley, 2008; Mesoudi, 2011). Not all stochastic processes need appear as a diffusion process (Der et al., 2012, 2011), yet at least superficial consistency is often observed (Bentley et al., 2011). In practice, if a more nuanced model is not necessary to achieve predictive or explanatory power over cultural change at a population level, then there may be no need to resort to one. An ongoing challenge of cultural evolution is then to determine whether reasoning in terms of the diffusion process can make reliable predictions, either about the future behavior of a population or about some properties of underlying mechanisms of transmission, even when the detailed nature of the cultural replicators is not known.

The work in Chapter 2, then, furthers this empirical project by using the neutral model—a diffusion process with zero mean change—as a null model to account for observed changes. When the null model can be rejected, it points to either a diffusion process that has a non-zero mean change, or potentially what is not a diffusion process at all. I rationalize the observations of a non-zero mean change, where they occur, in terms of prior expectations for biased transmission of language elements based on the cognition of speakers, and hence this observation redounds to the credibility of the diffusion process as a good model of cultural change. Furthermore, fitting diffusion models to the data allows the partitioning of a change into a stochastic part (the variance), which I call stochastic drift, and a deterministic part (the mean), which I generally call selection[1]. By comparing the strengths of drift and selection across many observations of similar changes (e.g. the regularization of many different past-tense verbs over time), I begin to tease apart what mechanisms contribute to one or the other. Working in the other direction, it is possible to propose mechanisms of learning and other cultural contagion (e.g. Ferdinand and Zuidema, 2009; Reali and

---

[1]I call this part "selection" for simplicity and mathematical analogy, despite that selection in the sense Darwin used it (the deaths or failures of reproduction of unfit individuals by analogy with the culling employed by plant and animal breeders) has no real interpretation in cultural information, and despite that "selection" (in the sense used here) may represent the result of many different mechanisms that may have no analogs in evolutionary biology or indeed mechanisms that have clear analogs that go under different names, such as biased mutation.

Griffiths, 2010; Ferdinand et al., 2013), and ask how they could or could not contribute to population-scale selection and drift.

In Chapters 1 and 3, I extend the simple idea underlying the thesis to social evolution. Social evolution studies the selection pressures imposed by the social environment, that is, the aspects of an organism's environment that is determined by interaction with conspecifics. The fact that most interesting traits under social selection also themselves affect the social environment greatly complicates their study. The population-genetic neutral model is of little use here as a null model or otherwise, because the mere fact that selection pressures from the social environment are operating presumes that the traits are not neutral. Much as the Lotka-Volterra dynamics of n types might contain n^2 different interaction strength parameters, different strategies in a population-wide iterated prisoner's dilemma (Axelrod et al., 1987) or paper-rock-scissors (Kerr et al., 2002) have idiosyncratic interactions with other strategies, creating a complex form of selection that depends on the frequencies of all types, a potentially infinite set of parameters. Again, chaotic dynamics and unmeasurable parameters plague the space of possible models of social evolution. If the neutral model is too simple, and the general model too complex, where do we turn?

The restriction of demographic equivalence between types carves out an interesting space of tractable models that allow social selective pressures, which I advance throughout the thesis as a null model for social evolution. I intend the words "null model" in a broad sense. These are a space of models which describe pervasive phenomena yet which can be well understood and well fit to data. As "null models", I advance that they merit study in their own right in order to understand a base case from which more complex situations may deviate: not because they will support a particular formal statistical hypothesis test in a frequentist framework, but because their continued study should well produce a font of such tests as well as better intuition about the basic behavior of selection in a social context. For example, based on what I develop here, one could easily imagine a test for whether any type has an idiosyncratic relationship to another type. This framing as an alternative

hypothesis to the null model that all types are equivalent is a restatement of a fundamental question in biodiversity theory. But we will also see that from this "null model" emerges a basic conformity/anti-conformity axis which must be, to some degree, a property of many socially evolving systems, and the null model provides an explicit basis for understanding the extent to which frequency-dependent selection influences biodiversity.

I call the models which follow the restriction of demographic equivalence "exchangeable models" following an old use of the word exchangeable from probability theory (Diaconis and Freedman, 1980, due to de Finetti) subsequently applied in population genetics (Cannings, 1974; Ewens, 2004). There is a connection between the exhangeable models, frequency-dependent selection, and the demographic equivalence described above in the context of ecology and evolution. The connection is a near-equivalence, the full details of which would be tedious to work out. In broad outline, however, exchangeability forbids fitness functions that depend on type identity and thus relationships between identified types. If we restrict our attention to Wright-Fisher models that assign fitness to each type by whatever means, then there are broadly two cases left over. One case is exactly the frequency-dependent selection described in evolution textbooks (Futuyma, 2009), where fitness depends on frequency, and linear dependence and equivalent dependence across all types are implicitly assumed. The freqeuncy-dependent case of exchangeability is more restrictive than the frequency-dependence of some social evolution models of population-scale versions of paper-rock-scissors or the prisonner's dillema in which fitness may depend on the frequencies of any type, but it is general enough to describe all cases of frequency-dependence that occur in e.g. Futuyma (2009). The remaining case is exchangeable models in which fitness depends not on type identity but may still depend on particular relationships between types that do not depend on their identity. An example of such an exchangeable process is one in which fitness depends on rank, for example assigning fitness $w_1 = 1$ to the most frequent type, and $w_n = 1 - ns$ to the nth most frequent type. Other examples might include a model where the fitness of a type at count 2 is $1 + s$ if there is another type with count greater than 500, and $1 - s$ otherwise. One can imagine ways to make

9

such a process Turing-complete, or mix it with regimes of frequency-dependence to create an ugly compendium of monsters, most of philosophical origin with no clear examples in nature. These models are still exchangeable in the sense of probability theory, but suffer a pathology that leads me to exclude them from consideration. A better connection would be an exact equivalence between frequency-dependent models (suitably defined) and the non-pathological exchangeable models.

The connection extends to the unified neutral theory of biodiversity, which describes a model in which populations of a number of species undergo the same birth-death process. This is a proposed neutral model for biodiversity ecology and has been recognized as an example for how neutral models can be employed in eg cultural evolution (Bentley et al., 2004; Valverde and Solé, 2015). Ignoring any migration between island and metapopulations, the expected number of species ϕ_n in a (meta)population containing n individuals is, in terms of the total number of species S_M, the birth and death rates of individuals b and d and a rate of creation of new species ν:

$$\langle \phi_n \rangle = \frac{S_M P_0 \nu}{nb} \left(\frac{b}{d} \right)^n \tag{1}$$

Here P_0 is a normalizing constant on the probability P_n of observing a species with n individuals, which is $P_0 \prod_{i=0}^{n-1} bi/d(i+1)$ or equivalently $(P_0\nu/dn)(b/d)^{n-1}$. The probability distribution is consistent with Fisher's log-series: the number of species at abundance n is $S_n = \alpha x^n/n$ (where $x = b/d$). This is a kind of abundance spectrum: How many will be present at a given volume n. This is called the 'distribution of relative species abundance (RSA)'. The connection allows us to import this result as a statement about the distribution of type frequencies in the neutral model of evolution. Indeed, Volkov et al. (2003) cite the Ewens sampling formula, developed in population genetics, as (Ewens, 1972; Karlin and McGregor, 1972) in the context of biodiversity theory. The connection also runs in reverse: results for exchangeable models will apply to an "exchangeable" generalization of neutral biodiversity theory. I discuss RSA in the exchangeable model further in Section 0.2.

I have found great interest in the evolution of coordination on a common phenotype. This

is a case of positive frequency-dependence, and it captures properties of standardization in colloquial and technical senses. Standardization is a well-studied topic in economics (Katz and Shapiro, 1985), sociology, linguistics and biology (Wagner et al., 2016). Examples range as far as bamboo flowering period (Janzen, 1976), the genetic code (Knight et al., 2001), and programming languages (Valverde and Solé, 2015). I discuss some below. Coordination is most interesting when it depends on very little extrinsic selection—the ordinary type of selection due to the physical environment or other organisms, which could adequately be called *ecological*—and rather depends primarily on selection arising from interactions with members of the same species, which I call *social*. Once a standard has emerged, rare traits are selected against, which suggests that standards change infrequently. Standards thus often become reliable features of the environment, and social pressure to maintain them may persist beyond the individuals or even species that originated them. These reliable features of the environment then easily play roles in niche construction, and other factors influencing biodiversity (by partitioning populations) and evolvability—as in a manner evocative of Waddington's description of canalization in development (Waddington, 1942) or in fostering modular design (Wagner et al., 2016).

The literature on standardization in economics, linguistics, and technology provides a rich source of inspiration, results, and potential applications and interdisciplinary collaboration. A precise understanding of the generalities of standardization would explain patterns across a disparate collection of systems. Many concepts and definitions of standards exist that emphasize different aspects. Technologically-focused definitions of standards emphasize the role of standards as reliable features of the environment. Tassey (2000) of the National Institute of Standards and Technology writes: "A standard can be defined generally as a construct that results from reasoned, collective choice and enables agreement on solutions of recurrent problems." Wagner et al. (2016) defines a technological standard as "a set of specifications to which all elements of a product, process, or format conform." Those concerned with technological standards often distinguish between *de facto* and "promulgated" (Tassey, 2000) or "*de jure*" (Wagner et al., 2016) standards, according to whether standardization

arose spontaneously—as through an evolutionary process of copying via immitation— or through a concerted process of social consensus for which an evolutionary model would be at pains to account. Wagner et al. (2016) conclude that certain properties of standards are essential foundations for innovation: the use of few, specifiable components; compatible interfaces; and the composability of different components into combinatorially-many potentially useful objects. For these purposes, Wagner et al. (2016) consider standards to be any omnipresent or near-omnipresent features of a technical system, regardless of their origins. For example, they consider the regularization of the form and manufacturing processes of stone tools in the Palaeolithic to be standardization regardless of contention as to whether the conformity is due to some form of social pressure or convergence towards an optimum design.

Chapter 1 investigates a very particular special case of standardization of some arbitrary traits. There, I focus on investigating the behavior of positive frequency dependent selection in a case that cannot be quite described in terms of the classical positive frequency dependence of underdominant traits. The project is to understand, in the details, how a general model of frequency-dependent selection can approximate a real system when many complex mechanisms such as diploidy, Mendelian inheritance, an mate choice are at play. Most of the value of this case study, sadly, can only be appreciated in the context of this thesis, and thus, on its own as it is published in *Theoretical Population Biology*, it amounts to a rather boring paper. What give it life and purpose in the context of the larger body of work are the realizations that many forms of frequency-dependence are possible outside those that have been studied classically, that the exact form of frequency-dependence can be deduced mathematically from a complicated set of mechanisms, and that alternative forms of frequency-dependence do bear on the resulting dynamics of the system. Most importantly, although this is not described in the paper, it demonstrates that a complicated model nonetheless reduces, via the one-dimensional manifold that is described in the paper, to a form of frequency-dependence.

Chapter 3 infers that form of frequency-dependence from timeseries data documenting the outcomes of competition between exchangeable types. This can provide a measure of standardization pressure, as we show by studying the case of human languages, or generally any kind of exchangeable frequency-dependent selection. In contrast to Chapter 1, it need no apology here, since I believe its merits will be evident outside the context of the thesis.

0.1. Cases of standardization

To study standardization, I collected case studies to use as thought experiments or sounding boards for the theory I present. Each case can be used to suppose whether a particular natural case of standardization fits the assumptions of the exchangeable model, to clarify how to interpret a real situation in the model, or to explore the diversity of mechanisms of frequency-dependent fitness.

The genetic code itself is a standard that is enforced by coevolution between sequences that encode protein products and the ribosomal machinery that creates the protein. Acridine mutagens can add or delete bases from the nucleotide sequence of bacteriophage T4. From little more than the phenotypes of recombinations of these mutations, Crick et al. (1961) determined that the genetic code consisted of non-overlapping triplets of nucleotides interpreted in order from a particular starting point. Today, the base of biophysical knowledge about the mechanisms of translation, the ribosome, and the genetic code is immense, yet the evolutionary origins of the ribosome and the genetic code remain mysterious.

A large literature deals with the evolution of the genetic code (Knight et al., 2001). The dominat view is that the universal genetic code arose by a frozen accident (Crick, 1968): once the genetic code is established, any alternative versions of the ribosome, tRNAs, aminoacyl-tRNA synthetase, or release factors that would alter the genetic code would cause mistranscription of any proteins containing the reassigned codon. Such pleiotropy would almost certainly cause deleterious effects. However, variant genetic codes do exist in some organisms, particularly organellar genomes or in intracellular parasites. Many theories have been

proposed to explain how alternative codes might arise against the initial deleterious effects of alternative interpretations of the existing sequence data (Knight et al., 2001).

The genetic code could be construed, on long timescales, as a coordination between genes. The optimal behavior at each codon is to generate the distribution of gene products that optimizes their functionality. Each gene would prefer one interpretation of a codon over another, to the degree that the alternative gene products differ in fitness. Where the preferred mapping is monomorphic across genes, then the optimal behavior is to generate the preferred target with high fidelity. When such a monomorphism is absent, as when genes horizontally transfer from a different system (eg nuclear to plastid gene transfer) or when tRNA mutations compensate for lethal non-synonymous substitutions (Murgola, 1985), the optimal behavior may be to stochastically generate alternative targets. This situation is similar to the ambiguous intermediate hypothesis of Schultz and Yarus (1994, 1996). Assuming that the targets and the specificity of each codon are evolvable—and there are several mutational mechanisms available between the tRNA anticodons that bind the code, the tRNA structure, and the synthetase that loads the amino acid onto the tRNA—we could assume that the ribosome, under high selection pressure for optimizing the aggregate of its gene products, evolves quickly. The genes, evolving on a slower timescale, evolve to coordinate their compliance with the current optimal codon behavior.

This coordination game results in standardization. Genes across the genome in the limit of weak mutation form a population in the standards process. Each locus bears a standard, ie its own idiosycratic preference for which amino acids it would like codons to represent. We assume the translation machinery generates amino acids for a codon (potentially ambiguously) so as to optimize the aggregate fitness of all loci sharing this codon. The translation machinery is then likely to produce the highest-fitness amino-acid at those loci bearing the dominant standard (amino-acid preference), and unlikely to produce the highest-fitness possible amino acid for loci still hanging on to rare standards. Under weak mutation, new mutations that bring loci into conformance with the dominant standard—nucleotide substi-

tutions that convert a codon to one more likely to produce a more-preferred amino acid—are more likely to fix than mutations to rare standards.

In this situation, the set of possible standards is finite, but nearly inexhaustible, and the exact Markov process depends on the distribution of fitness effects of amino acid substitutions and the mutational accessibility of different standards. Considering the exact (non-neutral) standards process for some set of genes, where the fitness effects are known, ie through simulation, might show that some structures that relate mutational accessibility to biophysical properties of amino acids will evolve in any genetic code. Relationships like these have been demonstrated (Freeland and Hurst, 1998). The standards process, by contrast, would provide a dynamic mechanism that explains the early evolution of the genetic code. This in turn would tease apart which features of the genetic code, such as the statistical overabundance of synonymous nucleotide substitutions, served an adaptive purpose (such as efficiency at minimizing the effects of errors, as has been assumed) and which were the result of dynamic constraint. Describing genetic code evolution with an exchangeable standards process could also relate some average fitness effects of amino acid substitution to mistranslation rates across the genome, a kind of mutation-selection balance against alternative standards (considered equivalent).

The order of genes on chromosomes is also an important standard due to the mechanisms involved in meiosis. Considerable work has gone into investigating the fixation rate of underdominant mutations due to a long controversy (White, 1978; King, 1995; Grant, 1991; Nevo et al., 2001; Walsh, 1982) about the involvement of chromosomal rearrangements in speciation (Noor et al., 2001; Rieseberg, 2001; Navarro and Barton, 2003b,a). Chromosomal rearrangements are classic underdominant mutations, as they often have no phenotypic effect, but cause some degree of hybrid sterility, a fitness detriment to heterozygotes. Lande (Lande, 1979) developed Kimura's work (Kimura, 1962) into an analytical expression for the fixation rate of a chromosomal rearrangement, which applies more generally to any

symmetrically underdominant alleles in diploid populations:

$$u = e^{-N_e s} \sqrt{\frac{s}{N_e \pi}}, \tag{2}$$

where u is the probability of fixation from initial introduction, N_e is effective population size, and s is the proportionate heterozygote disadvantage. The equation is valid when $N_e s > 2$. In the language of the standards model, symmetrically underdominant alleles are equivalent to standards with linear frequency dependence. On the basis of this formula and observations of rearrangements, Lande estimated local effective deme sizes in the tens or hundreds of individuals in organisms such as *Drosophila*, an inspiration for the work in Chapter 1. Although *Drosophila* escape this paradox in practice by suppressing recombination within inversions (and thus mitigating the partial sterility phenotype), mammals are subject to it. Human populations demonstrate a high selection pressure toward standardization. Humans show very little inversion polymorphism, and vanishingly little polymorphism for other chromosomal rearrangements. Known inversions are easy to detect with primers targeted towards the inversion endpoints. However, the discovery of unknown inversions requires reads across inversion breakpoints, necessitating good coverage of the entire genome. Researchers have used indirect methods (Bansal et al., 2007) for detecting inversion polymorphism in humans. There is some direct evidence that small inversion (up to 1Mb) polymorphisms exist in human populations (Stefansson et al., 2005; Feuk et al., 2005). This leaves a puzzle as to why many closely-related mammal species (such as ungulates) differ widely in their genetic architecture.

Human Language is an important standard and a topic of Chapter 3. In an increasingly globalized society, rare languages are going extinct. Abrams and Strogatz (2003) analyze the usage frequency of disappearing languages, fitting the transition probability to a language as a power-law of its frequency $P_{yx}(x, s) = cx^{\alpha}s$, $P_{xy}(x, s) = c(1-x)^{\alpha}(1-s)$, they inferred $\alpha = 1.21 \pm 0.25$ across 42 languages.

Words within languages are also the subject of population-scale coordination. Pagel et al.

16

(2007) and Fontanari and Perlovsky (2004) discuss how a neutral infinite-alleles model can result in power-law distributed word frequencies. Pagel alludes to evidence "that when more than one word is used to express the same meaning, the relative frequencies of use of the rare words is lower than expected from a neutral drift model of evolution (Fontanari and Perlovsky, 2004), consistent with selection against innovations."

Programming Languages are an analogous case in technological culture. Despite decades of software development, the barriers to compatibility between software written in different programming languages remain high. Thus, programming languages deliniate internally-cohesive bodies of work or functionality. The more functionality the collected works in a given language express, the more attractive that language is for new work. Valverde and Solé (2015) model this system of "increasing returns" and draw an explicit comparision (via the model) between programming language use and VHS versus Betamax, and a metaphorical comparison with the neutral theory of biodiversity. There are several indices (eg. TIOBE, PYPL, Transparent Language Popularity Index) that employ the number of results in search engines or popularity of programming languages in internet speech. Some (eg RedMonk) count github projects or stack overflow questions. These afford ready datasets to measure the standardization pressure of programming language use in the software engineering community.

0.2. Properties of the exchangeable models

Part of what makes the exchangeable models worth investigating is that some aspects of their behavior can be determined analytically. Many existing results from population genetics also have bearing on the exchangeable models, particularly from diffusion theory. Some results hinge on whether detailed balance holds in the exchangeable models, and a partial investigation of this question is relegated to the Appendix.

Kimura (1962) adapted the Kolmogorov backward equation to the case of allele frequency change under zygotic selection and random mating. In the general form of the Kolmogorov

equation, the fixation probability as a function of initial frequency $u(p)$ is (Kimura, 1962)

$$u(p) = \frac{\int_0^p G(x)dx}{\int_0^1 G(x)dx} \; , \; G(x) = e^{-\int \frac{2M_{\delta x}(x)}{V_{\delta x}(x)} dx} \tag{3}$$

where $M_{\delta x}(x)$ and $V_{\delta x}(x)$ are the mean and variance of the change in allele frequency per generation starting at allele frequency x. This is based on a continuum (diffusion) approximation of large population size. These equations can be specialized to arbitrary frequency-dependence of the marginal fitness of alleles by substitution of the appropriate $M_{\delta x}(x)$ and $V_{\delta x}(x)$. For example, in genic selection, $M_{\delta x}$ and $V_{\delta x}$ take the simple forms $sx(1-x)$ and $\frac{1}{2N}x(1-x)$ respectively. In the exchangeable models, $M_{\delta x}$ takes the form $s(x)x(1-x)$ where now s is viewed as a function of x.

Diffusion theory provides a way to study the equilibrium number of segregating sites (that is, polymorphic sites, existing in more than one variant or allele). Ewens (2004) furnishes an explanation of the long-term average frequency of an allele when two alleles can mutate to each other at a single site, using the same parameterization of Equation 3 as Kimura (1962). Watterson (1975) provides a more general result for the distribution of number of alleles present in a sample under neutral dynamics of infinitely many alleles. For example, the number of alleles present in the population in this model is

$$4N_e\mu \left[1 + \int_{1/2N}^1 x^{-1}(1-x)^{4N_e\mu-1} dx\right]. \tag{4}$$

Wright (1969) and Ewens et al. (1969) also derive equivalent expressions for this mean.

Frequency distributions of types have been of interest in population genetics. The Ewens sampling formula associates a probability with every partition of a sample (Ewens, 2004) from a neutral Wright-Fisher process at equilibrium, also called the multivariate Ewens distribution (Johnson et al., 1997) at stationarity:

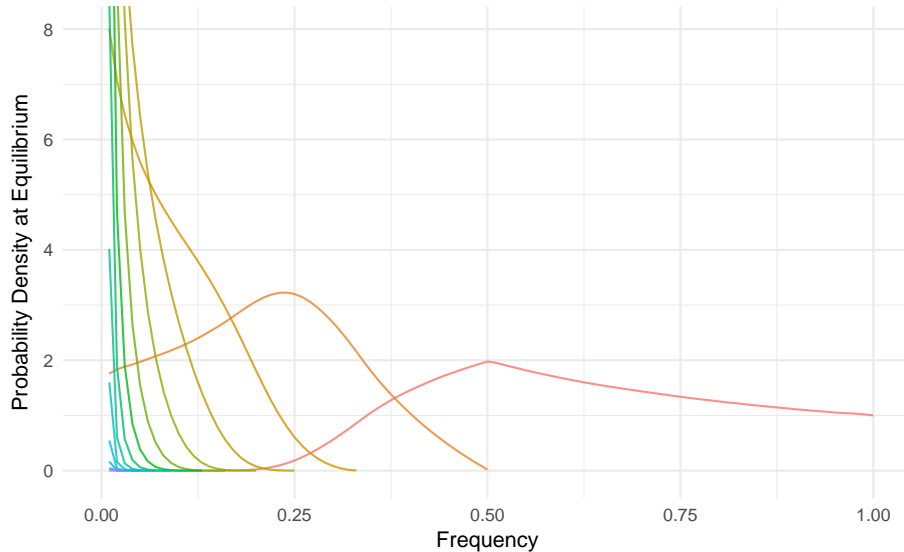$$\Pr(A_1, ..., A_N = a_1, ..., a_N) = \frac{N!\theta^{\sum_i^N a_i}}{\prod_i^N i^{a_i}a_i!(\theta+i-1)} \tag{5}$$

Figure 1: Rank abundance distribution of the nth most abundant type in an infinite-alleles Moran process with a population size of 100 and mutation to a new type a rate of once per 99 births (ie $\theta \approx 1$), by simulation after 100,000,000 births. The mean frequency of the most abundant type $(X_{(1)})$ is 0.624 according to Watterson and Guess (1977) and 0.628 in our data. Griffiths (1979) provide an analytical expression for the expected frequency of the ith most frequent type, and enumerate the mean frequency of the three most abundant types as $((X_{(1)}), (X_{(2)}), (X_{(3)})) = (0.624, 0.210, 0.088)$. Our data yields (0.628 0.210 0.088). Griffiths (1979) likewise enumerate the variance in frequency of the three most abundant types as $((X_{(1)}), (X_{(2)}), (X_{(3)})) = (0.037, 0.012, 0.005)$. In the above data, this is (0.037, 0.013, 0.005).

Ewens develops the sampling formula recursively using the probability of common ancestry within a sample, under the assumption that the sample size n is much less than the population size N (Ewens, 2004, p.114). Ewens expresses the formula as a probability distribution on a vector $A_{j \in \{1...n\}}$, $\sum_j j A_j = n$, of the number of allelic types present j times in the sample. Furthermore, Ewens quotes (due to Trajstman 1974), that the same formula for the probability is exactly true for the full population in the Moran process (p. 118).

In the exchangeable models, the distribution of states in the Markov process does not converge to an equilibrial value, as new types are constantly entering and leaving the population. Nonetheless, rank-abundance distributions do converge, since ranks are de-labled configurations. Empirical distributions of the neutral frequency distributions can be seen
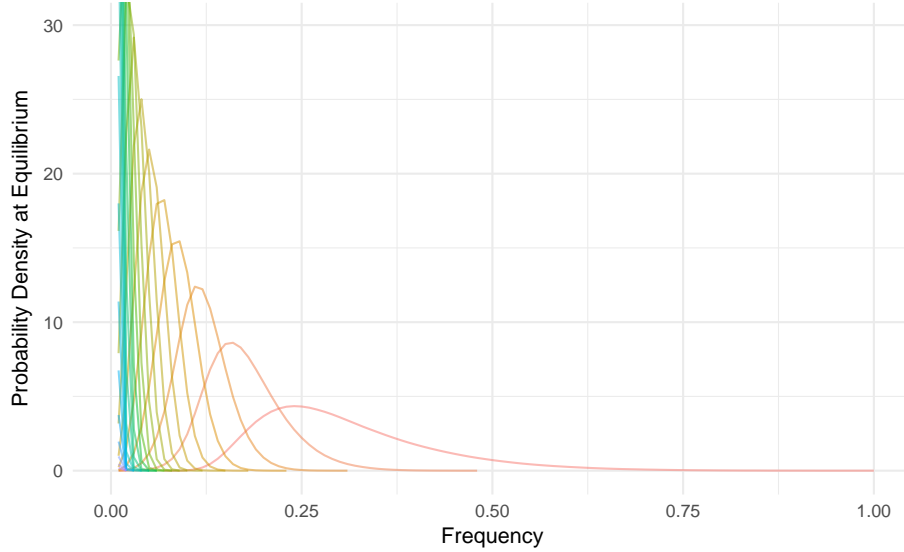
19

Figure 2: Rank abundance distribution with a population size of 100 and mutation to a new type a rate of 0.99*5 per 99 births $\approx$ 5 per haplotype per generation. Griffiths (1979) enumerate the mean abudance of the three most abundant types as $((X_{(1)}), (X_{(2)}), (X_{(3)}))$ = (0.34, 0.14, 0.04). Our data yields (0.307 0.175 0.119)

in Figs. 1 and 2. In determining whether the most frequent allele in a neutral population is also the oldest, Watterson and Guess (1977) determined the equilibrium frequency of the most frequent allele in an infinite-allele diffusion model. Griffiths (1979) extended this result to the expectation and variance of the frequency (at equilibrium) of the rth most frequent allele. Griffiths also includes allele frequencies in symmetric overdominance. The joint probabilities of delabeled configurations $X_{(i)}$ are given (albeit in a somewhat convoluted way) in the Moran process by Eq. 5.

We would like the distribution of the ith most abundant type at equilibrium, and we have the distribution of counts of all types. There is a 1-to-1 map between the vector A and any other representation of the partitions of n, though the map itself may be unwieldy. For example, we would like to order the abundance of alleles present in the population $x_1, x_2, ..., x_n$, such that $x_1 \geq x_2 \geq ... \geq x_n$. Transforming from the ordered abundance representation to the A representation is straightforward: $a_j(x) = \sum_{i=1}^{n} \delta_{j,x_i}$. This allows us to write the probability of observing x as $\Pr(A = a(x))$. The inverse map $x(a)$ exists,

20

but is difficult to write down algebraically. However, a recusive algorithm is:

$$x(a) = u(\mathbf{0}, 1, a, 1, 0)$$

where $\mathbf{0}_i = 0, i \in \{1, ..., N\}$, and

$$u(x, i, a, N, 0) = x,$$

$$u(x, i, a, j, c) = \begin{cases} u(\{x : x_i = j\}, i+1, a, j, c-1) & c > 0 \\ u(x, i, a, j+1, a_{N-j-1}) & c = 0 \end{cases}.$$

We know that $Pr(X_1 = x_1) = \sum_{x_2+x_3+...+x_N=N-x_1} Pr(X_1, ..., X_N = x_1, ...x_N)$, where the sum is over $p(N - x_1)$ terms (here, $p(n)$ counts the partitions of the integer n). The inner probability is given by Eq. 5 composed with the function $a(x)$ which transforms representations as described above. The existence of the mappings suggests that computing the distribution of the ith most abundant type is possible. Naive approaches by summing over the partitions are computationally intractable, but analytic reductions of Eq. 5 across subsets of the summand may simplify the computation. This may be an avenue for further progress.

It is interesting that in mutation-dominated regimes, the rank-abundance distribution follows a power law, as seen in Fig. 3. Bentley et al. (2004) measure power-law exponents in the rank-abundance distributions of cultural types (names, pottery motifs, patent citations) to infer the relative mutation rates $(N\mu)$, citing the Kimura neutral diffusion model. This seems to rely on an empirical statement that the distribution of baby names follows a power law coupled with the assumption that the distribution of baby names follows a diffusion process (Hahn and Bentley, 2003). In our data, the same empirical result seems to be true when $\theta \gg 1$, which reproduces the results of Bentley et al. (2004, Fig 2.). However, deviations have also been measured (O'Dwyer and Kandler, 2017). Mesoudi and Lycett (2009) take the basic approach of Bentley et al. (2004) and extend the simulations allowing
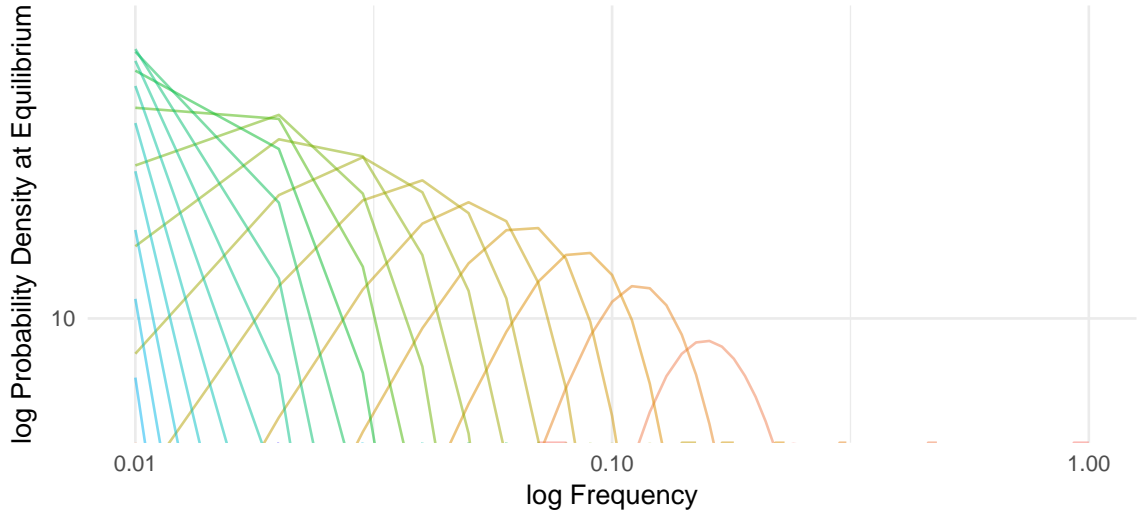
Figure 3: Data from Fig 2 log-transformed to show the approximate power-law distribution of expected frequencies.

copying processes that prefer or reject frequent or rare types. This affects the power-law distribution, which is not scale-free in general unless the allele dynamics are neutral with respect to frequency. This is a strange kind of process in which not frequency but rank matters to fitness—in the category of exchangeable but not strictly frequency-dependent process that I have excluded from my consideration. Connecting these observed power laws to the Ewens sampling formula (Eq. 5) is a goal of future research.

We have thusfar discussed models akin to the Wright-Fisher process, but other processes have been studied. Consider a pure-birth process, in which a pool of individuals exists, and each individual carries a trait. New individuals join the pool, and choose traints depending on the frequency of traints within the pool. As a motivating example, consider books. Books written in any of l languages are added to the library $L(t)$ with t books. We sort the library so that there are $L(t)_1$ books in the most common language, $L(t)_2$ in the second most common language, etc. Clearly $\sum_i^l L(t)_i = t$. We add a book in an existing language i with probability $f(L(t)_i/t)/Z$, and a new language with probability u. We could ask many reasonable questions. How does the long-term distribution of languages $L(t)/t$ in the library depend on a frequency-dependent growth rate? Does the frequency distribution of languages

22

converge to a particular value or does it fluctuate endlessly as a diffusion process? What is the distribution of end states? We could also ammend the pure-birth model to suppose that books are also lost from the library at random (irrespective of their language), or also allow the death rate of books to depend on their language. Would this affect the outcome?

Karlin and McGregor (1967) and Watterson (1974) study a continuous time process in which new 'mutant' indivisuals appear at a rate ν (regardless of the current population size), and birth and death of a type already present in the population both occur at a rate λ per individual of that type. The population size in this model eventually increases without bound due to the flux of new mutants. In this case, the distribution of A_j, the number of types represented by j individuals is known at any given time, and moreover, the A_j are independent. Watterson (1974) shows that this model has the same stationary distribution as a Moran model (described below) conditional on a certain population size. The equivalence between these two models makes some calculations easier, as in one model the population sizes are independent, see eg Griffiths (1979). Thus, the birth-death process bares a strong resemblence to the Wright-Fisher process in frequency space. This correspondence rationalizes somewhat a decision in Chapter 3 to model births of individuals within a given year as a population of names, rather than the total living individuals: The cases share a diffusion behavior in frequency space.

Spatial extensions are also possible. Likewise Valverde and Solé (2015) employ a lattice model, in which a programmer can adopt a programming language from adjacent cells on an $L \times L$ grid. Introducing space changes the variance structure of the change in frequencies, and thus is cannot be accomodated by only tracking a frequency state space. This is why, in investigating diffusion in frequency space, it is important to check empirically that the variance structure of the data is that of the Wright-Fisher diffusion (where the diffusion rate at frequecy x is proportional to $x(1 - x)$). This variance structure is evident in for example Fig. 10 of Chapter 2.

CHAPTER 1 : Positive Frequency-Dependent Fitness

Text from this chapter has been published in *Theoretical Population Biology* (Newberry et al., 2016).

Underdominant mutations have fixed between divergent species, yet classical models suggest that rare underdominant alleles are purged quickly except in small or subdivided populations. We predict that underdominant alleles that also influence mate choice, such as those affecting coloration patterns visible to mates and predators alike, can fix more readily. We analyze a mechanistic model of positive assortative mating in which individuals have n chances to sample compatible mates. This one-parameter model naturally spans random mating ($n = 1$) and complete assortment ($n \to \infty$), yet it produces sexual selection whose strength depends non-monotonically on n. This sexual selection interacts with viability selection to either inhibit or facilitate fixation. As mating opportunities increase, underdominant alleles fix as frequently as neutral mutations, even though sexual selection and underdominance independently each suppress rare alleles. This mechanism allows underdominant alleles to fix in large populations and illustrates how life history can affect evolutionary change.

1.1. Introduction

An allele is underdominant if it experiences reduced fitness as a heterozygote compared to either homozygote. Underdominance reduces diversity and has been studied as a mechanism for population differentiation and speciation (Wright, 1941; Lande, 1979). Underdominance typically occurs when the two homologous gene copies at a diploid locus must act in concert with one another. One classical example is a chromosomal alteration that disrupts meiosis in heterozygotes (Lande, 1979). Although much research on the fate of underdominant alleles has focused on chromosomal rearrangements, underdominant alleles arise and have evolutionary consequences in many other contexts. Underdominance has been observed at loci that regulate gene expression (Smith et al., 2011; Stewart et al., 2013), and engineered

underdominant transgenes play an important role in strategies to control insect disease vectors (Sinkins and Gould, 2006; Reeves et al., 2014). Underdominance has also been observed at loci controlling quantitative traits, such as body size (Kenney-Hunt et al., 2006), that are known to influence mate choice (Crespi, 1989). Likewise, heterozygote deficits in hybrid zones have been observed at the loci encoding color patterning in *Heliconius* butterfly species, where coloration has been implicated both in assortative mating (Jiggins et al., 1996; Arias et al., 2008) and also in viability via the avoidance of predators (Mallet and Barton, 1989; Kapan, 2001; Langham and Benkman, 2004). Motivated by these examples, we focus here on the fate of alleles that simultaneously influence mate choice and viability. We ask whether assortative mating will facilitate or impede the fixation of an underdominant allele.

In general, the fixation of an underdominant allele is exceedingly rare, at least in theory. A classical approximation due to Lande (1979) for the probability of fixation u of a novel underdominant allele with heterozygote disadvantage s in a well-mixed population of size N is:

$$u \approx \left(\frac{1}{N}\right) e^{-N_e s} \sqrt{\frac{N_e s}{\pi}}. \tag{1.1}$$

This fixation probability decreases exponentially with the effective population size, N_e. Under this analysis fixation through drift of a novel underdominant allele is possible only when the effective population size is extremely small—on the order of tens or hundreds of individuals. Nonetheless, empirical observations provide strong evidence that underdominant alleles have indeed fixed in populations.

Several possible resolutions to this paradox have been proposed. One common solution is based on Wright's shifting balance theory (Wright, 1931). According to Eq. 1.1, underdominant alleles may fix in extremely small populations, so that fixation across a species as a whole might occur through successive fixation in small, mostly-isolated subpopulations (Wright, 1941; Lande, 1979; Slatkin, 1981; Barton and Rouhani, 1991; Whitlock, 2003; Altrock et al., 2011). However, for this process to occur effectively it is typically necessary to include extinction and recolonization of demes (Lande, 1979; Michalakis and Olivieri, 1993;

25

Roze and Rousset, 2003). Other theoretically possible explanations (Hedrick, 1981) include meiotic drive and partial selfing (Charlesworth, 1992). Other authors have investigated linkage to locally adaptive alleles (Navarro and Barton, 2003a; Kirkpatrick and Barton, 2006).

Here we consider mate choice as a mechanism to explain the fixation of underdominant alleles at loci that influence both mate choice and viability. Phenotypes that serve as mating cues are often also subject to surveillance by predators, where rarity is typically detrimental to survivorship. In the context of coloration patterns in the *Heliconia*, for example, rare phenotypes often both experience increased predation and also determine assortment (Mallet and Barton, 1989; Kapan, 2001; Langham and Benkman, 2004; Jiggins et al., 1996; Arias et al., 2008). Likewise, in vertebrates ranging from cichlids (Sefc et al., 2014; Anderson et al., 2015) to finches (Blount, 2004; Blount et al., 2003), carotenoid coloration phenotypes are well known to influence mate choice and viability alike.

A broad literature has successfully addressed questions about diversity of mating systems in nature (Andersson, 1994; Andersson and Simmons, 2006; Clutton-Brock and McAuliffe, 2009), their evolutionary maintenance and optimality (Lande, 1981; Lande and Schemske, 1985; Real, 1990; Goodwillie et al., 2005; Kokko et al., 2006; Jones and Ratterman, 2009; Wiegmann et al., 2010), and their consequence for allele frequency change (e.g. inbreeding depression (Charlesworth, 1992; Nagylaki, 1992; Whitlock, 2000)). While there are many analytical studies on the effects of mating systems on allele dynamics, they tend to provide either a deterministic treatment under a specific model of mate choice (Karlin, 1978; O'Donald, 1980; Kirkpatrick, 1982; Seger, 1985; Otto et al., 2008), or a full stochastic treatment but only for mating systems that are essentially equivalent to a fixed population structure (i.e. a constant inbreeding coefficient, F (Caballero and Hill, 1992; Damgaard, 2000; Roze and Rousset, 2003; Glémin, 2012)).

Rather than stipulate a fixed population structure or a constant probability of selfing, we will provide a stochastic analysis of a mechanistic model of assortative mating. The model

is defined in terms of the absolute number of individuals, n, that an organism can survey before eventually choosing a mate. This one-parameter model of positive assortative mating coincides with classical partial self-fertilization in two limiting cases. For $n = 1$ the model corresponds to random mating, whereas as $n \to \infty$ it corresponds to complete assortment. Although the fixation probability of a new mutation under partial selfing smoothly interpolates between these two extreme cases (Charlesworth, 1992), we will show that the fixation probability in our model has a non-monotonic dependence on the life-history parameter, n. Increasing the number of mating opportunities beyond $n = 1$ initially inhibits the fixation of underdominant alleles, but increasing n yet further eventually facilitates fixation, allowing rates approaching that of neutral substitutions. These results are surprising because the mate choice model induces a form of positive frequency-dependent sexual selection that, in the absence of viability selection, always inhibits the fixation of rare alleles. We will explain these results in terms of the geometry of a slow manifold that arises under preferential mating, analogous to the Hardy-Weinberg equilibrium for random mating, and we discuss implications for the evolution of underdominant alleles in nature.

1.2. A mechanistic model of mate choice

Models of mate choice, assortative mating, and sexual selection have been extensively studied and characterized (Gavrilets, 2004). In partial self-fertilization or mixed mating models, individuals mate with themselves with a fixed probability and the mating system does not alter allele frequencies (Haldane, 1924). In partial assortative or preferential mating models, parents prefer to mate with their own genotype, or with particular other genotypes, and the mating system itself can alter allele frequencies by sexual selection. Such models are typically formulated (Karlin, 1978) by specifying, exogenously, the chance that one genotype will mate with another genotype, which allows mating with rare types according to preference regardless of the frequency of the rare type. This formulation implies that individuals are able to census all other individuals in the entire population in the decision to mate—which is unrealistic for many biological populations. Even some models that in-

corporate a search cost (Otto et al., 2008) still have the property that the cost of finding vanishingly rare types is fixed, regardless of their frequency.

Here we study a one-locus, two-allele model of hermaphroditic diploid individuals in which parents prefer to mate with their own genotype. This models situations where each genotype has a distinct, visible phenotype subject to assortative mating, and viability selection disfavors the heterozygous phenotype. Over the course of successive discrete generations we track the frequencies of all three diploid genotypes, x_i for $i \in \{aa, aA, AA\}$, in a population of constant size N. A parent can mate with any individual from a pool of n prospective mates, drawn uniformly with replacement from the population. If there is a mate of the parent's own genotype among these n prospective mates, then the parent chooses that mate; otherwise, the parent chooses uniformly at random from the pool of n prospective mates. The mating always produces one offspring. This model is equivalently described as parents having n chances to find a compatible mate by sampling randomly from the population. Parents sample mates from the population up to n times, mating immediately with any individual of their own genotype, or ultimately accepting any genotype on their nth chance. Because all parents choose among n potential partners, we call this the n-choice model. Several similar models, proposed by O'Donald (1980) and Janetos (1980), have previously been analyzed in a deterministic setting (O'Donald, 1980; Seger, 1985).

The n-choice model is a simple, mechanistic implementation of positive assortative mating that accounts for the fact that in reality rare genotypes are less likely to find preferred mates. Importantly, the model does not rely on an intrinsic capacity for selfing. In other words, there is no fixed chance that an individual will reproduce with its own genotype regardless of the frequency of that genotype. Rather, individuals census a limited number of possible mates from the population with replacement. While this sampling scheme includes a small probability that an individual will 'encounter' and mate with itself, we demonstrate robustness of the results to this and other biological assumptions in the Discussion section.

The outcomes of mating are dependent on all genotype frequencies, and so we must tabulate

the probability of each mating (Nagylaki, 1992). The n-choice model does not explicitly distinguish between sexes, as with hermaphroditic or monoecious populations. Nevertheless, we may consider the mate-choosing parent in any pairing as the "female" or macrogamete-donor parent. According to the model, the probability that a female (i.e. mate-choosing) parent with diploid genotype i finds a mate of her own genotype is $(1 - (1 - x_i)^n)$. Otherwise, the parent reproduces with a different genotype with probability proportional to that genotype's frequency in the population. The probability distribution of mating types G conditional on the female's genotype P is thus

$$\Pr(G = j \times i | P = i) = \begin{cases} 1 - (1 - x_i)^n, & i = j \\ x_j(1 - x_i)^{n-1}, & i \neq j \end{cases}, \text{ for } i, j \in \{aa, aA, AA\}.$$

The genotypic distribution of offspring from a given mating pair is Mendelian. We can therefore compute the distribution of zygotic genotypes Z after mate choice and reproduction by conditioning on the distribution of matings:

$$\Pr(Z = k) = \sum_i \sum_j \Pr(Z = k | G = i \times j) \Pr(G = i \times j | P = i) x_i \tag{1.2}$$

where $\Pr(Z = k | G = i \times j)$ denote the standard Mendelian probabilities, and i, j, k range over the three diploid genotypes aa, aA, and AA.

Following mate choice and production of a large zygotic pool we assume that viability selection modifies the frequencies of zygotic genotypes. The subsequent generation of reproductive adults is then drawn from the post-selection zygotic pool. Assuming the zygote pool is very large relative to the population size, the genotype of each surviving sampled adult in the next generation is drawn from the trinomial, fitness-weighted zygote distribution.

Genotype i is drawn with probability

$$\frac{Pr(Z = i)w_i}{\sum_j Pr(Z = j)w_j} \tag{1.3}$$

We express the fitness scheme for underdominant alleles as $w_{aa} = w_{AA} = 1$, and $w_{aA} = 1-s$.

1.3. Analysis

We explore the influence of mate choice on the fixation rate of alleles by analyzing the n-choice assortative mating model in finite populations.

In a finite population of constant size N adults, the frequencies of the adult genotypes of the next generation are drawn from the trinomial distribution with the probability of genotype i given by Eq. 1.3. We denote the outcome of this trinomial draw, for each genotype $i \in \{aa, aA, AA\}$, by x'_i. In other words, the values x'_i denote the frequencies of the three genotypes in the next generation of adults, given the frequencies x_i in the current generation.

The expected frequency of genotype i among adults in the next generation is given simply by the frequency of that genotype in the post-selection zygotes, that is by Eq. 1.3 above, which we henceforth denote $\mathrm{E}(x'_i)$. Under the multinomial sampling assumption the variance in the frequency of genotype i among the adult individuals in the next generation is simply $\mathrm{E}(x'_i)(1 - \mathrm{E}(x'_i))/N$, and the covariance between different genotypes is $-\mathrm{E}(x'_i)\,\mathrm{E}(x'_j)/N$.

Genotype frequencies exist on the simplex $x_{aa} + x_{aA} + x_{AA} = 1, x_i > 0$. We can thus remove one variable from the model by tracking genotype frequencies in a two-dimensional basis. We will track the frequency of the a allele among adults, denoted $p = x_{aa} + (1/2)x_{aA}$, and one-half the frequency of heterozygous adults, denoted $h = x_{aA}/2$. One-half the frequency of heterozygotes may be thought of as the frequency of the a allele present in heterozygotes, and so it has the same units (allele copy frequency) as p. The coordinate h is preferable to other measures of heterozygosity, such as the quantity $x_{aA}/(p(1 - p))$, because it does not

depend on allele frequency p.

We can express the expected allele frequency $\mathrm{E}(p')$ and the expected heterozygous allele frequency $\mathrm{E}(h')$ in the next generation in terms of the current frequencies p and h. To do so, we first write the mean allele and heterozygote frequencies among zygotes prior to selection as π and η in terms of p, $q = 1 - p$ and h by expanding Eq. 1.2:

$$
\begin{aligned}
\pi = {} & h(1 - (1 - 2h)^n) + 2h(1 - 2h)^{n-1}\left(\frac{p}{2} - h + \frac{1}{4}\right) \\
& + \frac{1}{2}\left(q - h\right)\left(p - h\right)\left((p + h)^{n-1} + (q + h)^{n-1}\right) \\
& + 2h\left(\frac{3}{4}\left(p - h\right)\left(q + h\right)^{n-1} + \frac{1}{4}\left(q - h\right)\left(p + h\right)^{n-1}\right) \\
& + (p - h)\left(1 - (q + h)^n\right)
\end{aligned}
\tag{1.4a}
$$

$$
\begin{aligned}
\eta = {} & \frac{h}{2} + \frac{h}{2}\left((p - h)\left(q + h\right)^{n-1} + (q - h)\left(p + h\right)^{n-1}\right) \\
& + \frac{1}{2}\left(q - h\right)\left(p - h\right)\left((p + h)^{n-1} + (q + h)^{n-1}\right)
\end{aligned}
\tag{1.4b}
$$

We then use Eq. 1.3 to obtain expressions for the frequencies of post-selection zygotes in the (p, h) basis, which are the expected frequencies of adults in the next generation:

$$
\mathrm{E}(p') = \frac{\pi - s\eta}{1 - 2s\eta}, \qquad \mathrm{E}(h') = \frac{(1 - s)\eta}{1 - 2s\eta}.
\tag{1.5}
$$

Finally, the variance in adult allele frequencies in the next generation can then be written as

$$\mathrm{Var}(p') = \left(\frac{1}{N}\right)\left(\left(\mathrm{E}(p') - \mathrm{E}(h')\right)\left(\mathrm{E}(q') + \mathrm{E}(h')\right)\right. \tag{1.6}$$

$$\left. + \frac{1}{2}\,\mathrm{E}(h')(1 - 2\,\mathrm{E}(h')) + 2\,\mathrm{E}(h')\left(\mathrm{E}(p') - \mathrm{E}(h')\right)\right)$$

$$\mathrm{Var}(h') = \mathrm{E}(h')(1 - 2\,\mathrm{E}(h'))/2N$$

$$\mathrm{Cov}(p', h') = \mathrm{E}(h')(1 - 2\,\mathrm{E}(p'))/2N.$$

Here we see that $\mathrm{Cov}(p', h')$ is positive only if $\mathrm{E}(p') < 1/2$. This expectation and variance will be sufficient to develop a diffusion approximation to the model, along the lines of Kimura (1964).

1.3.1. Behavior in two classical limits

When the sampled pool of prospective mates has only one individual, $n = 1$, or when the pool becomes much larger than the population itself, $n \to \infty$, the n-choice mating model corresponds precisely to two classical population models: random mating and complete inbreeding, respectively.

When only one prospective mate is allowed per parent, $n = 1$, the model reduces to the classical model of an underdominant allele in a randomly mating population. Eqs. 1.4 simplify to $\pi = p$ and $\eta = p(1 - p)$. Thus the genotype frequencies among zygotes are at Hardy-Weinberg equilibrium, and viability selection proceeds as in ordinary underdominance with $\mathrm{E}(p') = (p - sp(1 - p))/(1 - 2sp(1 - p))$. When $s = 0$, the dynamics are neutral on $2N$ haplotypes. When there is selection against heterozygotes, $s > 0$, then the fixation probability of a novel allele agrees with Lande's classic expression, given by equation 1.1.

On the other hand, when the number of mating opportunities becomes very large, $n \to \infty$, preferred genotypes are always available for mating, and the model is equivalent to complete assortment. As we take n to infinity in Eq. 1.4, the frequency of heterozygotes among zygotes

approaches $\eta = h/2$. That is, the frequency of heterozygotes is reduced by half at each generation. The population thus rapidly approaches heterozygote frequency zero. Setting $h = 0$ in Eq. 1.4 and taking n to infinity gives $\pi = p$. As selection acts only on heterozygotes, $\mathrm{E}(p') = p$ and the dynamics are neutral. The variance of p' is $\mathrm{E}(p')(1 - \mathrm{E}(p'))/N$. This variance is the same as that for a population of N haplotypes, rather than the $2N$ actually present in the population. Because of complete assortment, each diploid individual behaves roughly as a single haplotype. In this limit of complete assortment, the dynamics of genotype frequencies are always neutral, regardless of s, and the fixation probability of a novel allele is always $1/2N$.

When the number of mating opportunities is intermediate between these two extremes, namely $1 < n < \infty$, the dynamics of the n-choice model are neither neutral nor the same as the dynamics of classical underdominance. Importantly, these dynamics may change allele frequencies dramatically due to a mix of sexual selection and underdominant selection. Analyzing these two-dimensional dynamics requires the development of an appropriate diffusion approximation.

1.3.2. Diffusion approximation

We adopt the techniques used to analyze the fixation probability of an underdominant allele under random mating to derive a more general expression for the n-choice mating model. Under the diffusion limit of Kimura (1964), the probability density $\phi(p, h, \tau)$ of observing allele frequencies p and h evolves in time according to the standard Kolmogorov forward equation (Gardiner, 2009), which depends on the instantaneous mean and variance-covariance matrix of the changes in allele frequencies.

To derive a diffusion approximation we start with the Kolmogorov forward equation in two dimensions (Gardiner 2009),

$$\frac{\partial \phi}{\partial t} = \frac{1}{2} \left(\frac{\partial^2 V_{pp} \phi}{\partial p^2} + 2 \frac{\partial^2 V_{ph} \phi}{\partial p \partial h} + \frac{\partial^2 V_{hh} \phi}{\partial h^2} \right) - \left(\frac{\partial M_p \phi}{\partial p} + \frac{\partial M_h \phi}{\partial h} \right). \tag{1.7}$$

Here, M_i and V_{ij} represent the instantaneous mean and variance of allele frequency change. Assuming only a fraction f of the population undergo mating according to Eq. 1.4, the expected frequencies in the next generation, p' and h', are given by replacing π and η with $f\pi + (1-f)p$ and $f\eta + (1-f)h$ in Eq. 1.5:

$$\mathrm{E}(p') = \frac{(f\pi + (1-f)p) - s(f\eta + (1-f)h)}{1 - 2s(f\eta + (1-f)h)}$$
$$\mathrm{E}(h') = \frac{(1-s)(f\eta + (1-f)h)}{1 - 2s(f\eta + (1-f)h)}$$

The mean change per generation is then $M_p = E(p') - p$ and $M_h = E(h') - h$, which are rational functions in p and h of height $2n$. We take the limit $N \to \infty$, while holding both $Nf = \zeta$ and $Ns = \gamma$ constant, so that f and s are small parameters. Thus only first-order terms survive in the Taylor expansion of M_p and M_h around $(f,s) = (0,0)$ even when Ns and Nf are large. We are left with

$$M_p = f(\pi - p) + sh\left(p - \frac{1}{2}\right), \quad M_h = f(\eta - h) - sh(1 - h). \tag{1.8}$$

As f and s approach zero in this limit, so does the mean change in allele frequency, and $\mathrm{E}(p') \to p$ and $\mathrm{E}(h') \to h$. Thus the variance-covariance matrix of allele frequency change approaches the multinomial variance-covariance matrix of sampling from the current allele frequencies. Thus the V_{ij} are given simply by Eq. 1.6: $V_{pp}(p,h) = \mathrm{Var}(p)$, $V_{hh}(p,h) = \mathrm{Var}(h)$, and $V_{ph}(p,h) = \mathrm{Cov}(p,h)$, where $E(p) = p$.

We rescale time in Eq. 1.7, taking $\tau = t/N$:

$$\frac{\partial \phi}{\partial \tau} = \frac{N}{2}\left(\frac{\partial^2 V_{pp}\phi}{\partial p^2} + 2\frac{\partial^2 V_{ph}\phi}{\partial p \partial h} + \frac{\partial^2 V_{hh}\phi}{\partial h^2}\right) - N\left(\frac{\partial M_p\phi}{\partial p} + \frac{\partial M_h\phi}{\partial h}\right). \tag{1.9}$$

Writing Eq. 1.9 in terms of π, p, η and h and combining factors $Nf = \zeta$ and $Ns = \gamma$ gives Eq. 1.10.

We compute the integral in Eq. 1.13 and the manifold of equilibrium heterozygosity (Eq. 1.11) numerically using Mathematica (Wolfram Research, Inc, Mathematica, Version 10.0.2.0 (2015), Champaign, IL, USA). We also wrote software in OCaml using the GNU Scientific Library to estimate fixation probabilities of the discrete model by explicit Monte Carlo simulation. The software is open source and available on GitHub (`https://github.com/mnewberry/nchoice`).

We introduced the parameter f, describing the proportion of the population that undergoes mate choice as opposed to clonal reproduction. Although this parameter was introduced for technical reasons, in order to produce a well-defined diffusion limit, even in finite models Nf has a natural, physical interpretation as the rate of mating: the average number of matings per generation, or the relative strength in altering gene frequencies by the mating system versus by genetic drift. One might naively assume that f is always unity in natural populations, and yet many plants such as some grasses and aspen reproduce sexually on a background of clonal reproduction. Genetic drift due to accidents of sampling can be interpreted at many levels, including sampling induced by the outcomes of mating; or stochastically induced by persistence to the next generation through longevity.

We find the instantaneous mean and variance of allele frequency changes, M_i and V_{ij}, by rescaling the discrete model by the population size, N. To arrive at a non-trivial diffusion limit we adopt a slight variant of the model above, in which only a fraction f of the population undergoes mate choice each generation, while the remainder of the population is sampled according to strict clonal reproduction. We take the limit as N approaches infinity, scaling f and s such that $Nf = \zeta$ and $Ns = \gamma$ are held constant. The diffusion equation becomes

$$\frac{\partial \phi}{\partial \tau} = \frac{\partial^2}{\partial p^2}\left((p-h)(q+h) + \frac{h(1-2h)}{2} + 2h(p-h)\right)\frac{\phi}{2} \qquad (1.10)$$

$$+ \frac{\partial^2}{\partial p \partial h}\frac{h(1-2p)}{2}\phi + \frac{\partial^2}{\partial h^2}\frac{h(1-2h)}{2}\frac{\phi}{2}$$

$$- \zeta\left(\frac{\partial}{\partial p}(\pi - p)\phi + \frac{\partial}{\partial h}(\eta - h)\phi\right)$$

$$- \gamma\left(\frac{\partial}{\partial p}h\left(p - \frac{1}{2}\right)\phi - \frac{\partial}{\partial h}h(1-h)\phi\right)$$

In general, this diffusion in allele-frequency space is two-dimensional and well known to be computationally formidable (Epstein and Mazzeo, 2013). In the next section we introduce a one-dimensional approximation that makes the diffusion tractable.

1.3.3. Diffusion along the slow manifold

The dynamics in two-dimensional diffusions sometimes approach and remain in the vicinity of a one-dimensional curve, until absorption into a boundary. This behavior can be interpreted as a separation of timescales (Parsons and Rogers, 2015): there is fast approach to a lower-dimensional manifold, and then slow diffusion along the "slow manifold". In the case of random mating for a single-locus diploid model, for example, Kimura's one-dimensional diffusion works because genotype frequencies are assumed to be at Hardy-Weinberg equilibrium at all times. The quasi-linkage equilibrium approximation in multilocus models (Kimura, 1965) is another example. For monoecious random mating, approach to the manifold of Hardy-Weinberg equilibrium takes only a single generation, which is instantaneous in the diffusion timescale. In other mating systems, such as random mating with separate sexes, equilibrium is reached after two generations of mating, or the slow manifold may be approached geometrically. Some dynamics, such as clonal reproduction, do not approach any lower-dimensional sub-manifold whatsoever, and so they exhibit truly two-dimensional diffusions.

Principled approaches to determining the existence and form of the slow manifold are com-
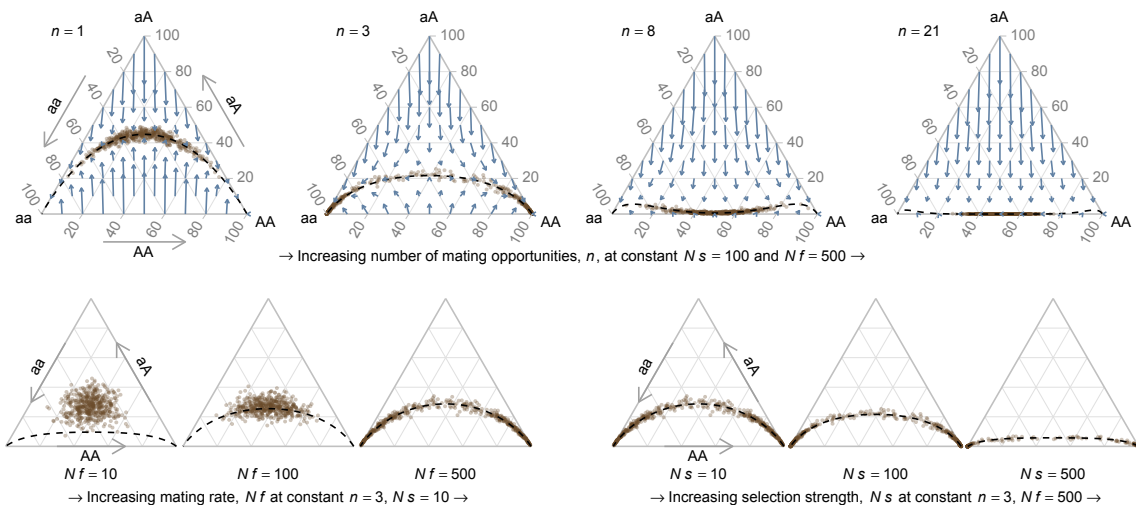
Figure 4: The dynamics of genotype frequencies in the n-choice model of assortative mating. Each ternary plot corresponds to a different set of parameter values for the number of mating opportunities, n, the per-generation rate of participation in the mating system, Nf, and the strength of viability selection against heterozygotes, Ns. Arrows indicate the expected change in genotype frequencies in one generation. On each plot, dots represent the genotype frequencies after 30 generations of stochastic simulation in 500 replicate populations of size $N = 1,000$ each initialized at the center ($\frac{1}{3}aa$, $\frac{1}{3}aA$, $\frac{1}{3}AA$). The dynamics quickly converge towards a one-dimensional submanifold within the frequency space. Dashed lines show the analytically-derived position of this one-dimensional manifold, which corresponds to the Hardy-Weinberg equilibrium in the case of random mating ($n = 1$). Increasing the strength of selection, Ns, moves the manifold towards zero heterozygosity, while the effect of participation in the mating system, Nf, on the shape of the manifold depends on the number of mating opportunities, n. In most regimes depicted, information about the initial height (heterozygosity) of the population is lost after 30 generations, as the genotype frequencies have converged to the slow manifold.

plex (Parsons and Rogers, 2015). Nonetheless, the dynamics of the n-choice mating model clearly exhibit timescale separation when viability selection or participation in the mating system are strong ($Ns \gg 10$ or $Nf \gg 10$), which can be seen in Fig. 4. When Nf and Ns are both weak, convergence to the manifold is slower than diffusion in genotype space due to clonal reproduction. However, when Ns and Nf are zero, the diffusion equation (1.10) correspond to classical clonal diffusion. In practice, we find the quality of the approximation good whenever $Nf > Ns$, at any value of n. We analyze the diffusion along this slow manifold, as opposed to other techniques, because of its simplicity and analogy to Hardy-Weinberg equilibrium.

To determine the manifold of equilibrium genotypic frequencies in the n-choice model we use the simple principle that frequencies at equilibrium should stay at equilibrium. Mathematically, this implies the condition that the infinitesimal mean change in frequencies (p, h) must always be tangent to the slow manifold. Thus, the slow manifold can be defined as a parametric curve $(p(l), h(l))$ such that

$$\frac{dp(l)}{dl} = M_p(p(l), h(l)), \qquad \frac{dh(l)}{dl} = M_h(p(l), h(l)). \tag{1.11}$$

This differential equation has an infinite family of solutions. Fortunately there are additional criteria for equilibrium genotypic frequencies. Since rare alleles are always present as heterozygotes, $dp(l)/dh(l)$ must approach 1 as l approaches infinity. This criterion in terms of long times is difficult to use in practice, and so we use the symmetry of the dynamics: when $p = 1/2$ the slow manifold is horizontal, meaning $E(h') = h$. We use this symmetry criterion to initialize the parametric curve close to $p = 1/2$. This particular solution to the differential equation above provides a function $\tilde{h}(p)$, which defines the manifold of equilibrium heterozygosity as a function of allele frequency.

If we assume that the two-dimensional dynamics in (p, h) are restricted to the slow manifold, defined by the curve $(p, \tilde{h}(p))$, then we can treat the dynamics as a one-dimensional diffusion along this manifold satisfying

$$\frac{\partial \phi}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial p^2}(V_{pp}\phi) - \frac{\partial}{\partial p}(M_p \phi), \tag{1.12}$$

by substituting $\tilde{h}(p)$ for h in Eq. 1.4. Although M_p and V_{pp} depend on h, we compute h and η from p by substituting $\tilde{h}(p)$ for h in Eq. 1.4, and we view M_p and V_{pp} as functions of p only. Following Kimura (1964), the solution to a boundary value problem gives the fixation probability $u(p)$ of a mutant allele initiated at frequency p with solution

$$u(p) = \int_0^p dx\, \psi(x) \bigg/ \int_0^1 dx\, \psi(x)\,, \ \psi(p) = e^{-\int_0^p dx\, [2M_p(x)/V_{pp}(x)]}. \tag{1.13}$$

This integral can be computed numerically, giving a good approximation to the fixation probability for arbitrary n.

1.4. Results

1.4.1. Sexual selection in the n-choice mating model

First we consider the fixation probability of an allele in the absence of viability selection. In the limits of random mating and complete assortment, $n = 1$ and $n \to \infty$ respectively, the fixation probabilities are equal to the neutral fixation probability. For an intermediate number of mating opportunities, n, however, the mating system itself induces strong sexual selection against rare alleles (cf. Fig. 4), which depresses the fixation probability of novel alleles much below the neutral probability (Fig. 5). And so the resulting fixation probability has a complex, non-monotonic dependence on the number of mates that a parent can survey: increasing n beyond one reduces the fixation probability below the neutral value $1/2N$, but increasing n yet further restores the fixation probability until it recovers to the neutral value $1/2N$ in the limit $n \to \infty$.

The strength of sexual selection against rare alleles depends on both the degree of participation in the mating system, Nf, and the number of mating opportunities, n. At high mating rates ($Nf \geq 100$), fixation probabilities at intermediate values of n are so low as to be impractical to compute through Monte-Carlo simulation, and they can be known only through numerical integration of the expression derived from the diffusion approximation, Eq. 1.13. In the regime where Monte-Carlo methods are feasible, the diffusion approximation along the slow manifold is in close correspondence with Monte Carlo simulations across a broad range of values of Nf, despite many potential sources of error in the approximation
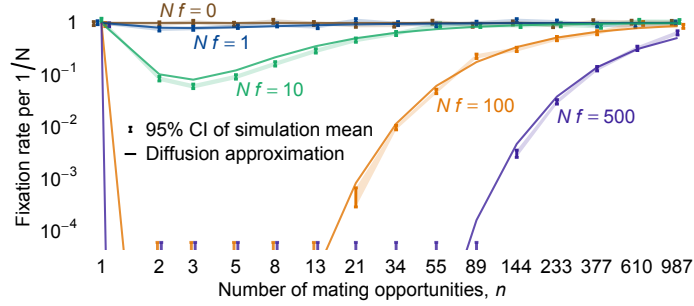
Figure 5: Sexual selection induced by the n-choice model of assortative mating. The plot shows the fixation probability of a new mutation (one initial heterozygote) in the absence of viability selection, as a function of the number of mating opportunities, n, and for different rates of participation in the mating system, Nf. Solid lines indicate fixation probabilities computed by the diffusion approximation along the slow manifold. Bands and error bars indicate the 95% confidence interval on the mean fixation rate observed in up to 100,000,000 simulated populations of size $N = 1,000$. Overlapping error bars are dodged for clarity. Error bars with no lower bound indicate no fixations observed in 100,000,000 simulations. The fixation probability equals the neutral fixation probability for either $n = 1$ or $n \to \infty$. Otherwise, participation in the mating system induces a complex form of sexual selection against rare alleles, whose strength depends non-monotonically on n.

(see Fig. 5).

When the number of potential mates n is small but exceeds one, then rare alleles are under-represented among zygotes relative to their parents, as described by Eq. 1.5. Mendelian inheritance does not alter allele frequencies, and so it is the action of mate choice itself that suppresses rare alleles: parents with common genotypes are likely to find their preferred mates, but parents with rare genotypes are more likely to settle for a common mate. As n increases, however, the likelihood of any parent having to settle for a non-preferred mate becomes vanishingly small, and so the differences in mate availability between genotypes become less pronounced. When there is no difference in mate availability even the rarest of parental genotypes can find a mate with their own genotype, and complete assortment ensues.

In summary, Fig. 5 reveals that a mechanistic model of mate choice with two classical limits nonetheless produces a complex form of sexual selection whose strength depends non-monotonically on number of mating opportunities, n.

40

1.4.2. Interaction between sexual selection and underdominant viability selection

What is the fate of novel alleles when we combine the intrinsic effects of n-choice mating with viability selection against heterozygotes? Does positive assortative mating facilitate or impede the fixation of an underdominant allele?

The combined effects of viability selection and sexual selection induced by n-choice mating are shown in Fig. 6. Here we see that when the number of mating opportunities is small but not one, then preferential mating impedes fixation of an underdominant allele, so that its fixation probability is even lower than the classical prediction of Lande for random mating ($n = 1$). For example, allowing just $n = 2$ opportunities to find a mate with the same genotype reduces the fixation probability of the underdominant allele astronomically.

Nevertheless, as the number of mating opportunities increases further we find a surprising result: viability selection against heterozygotes and n-choice assortative mating—two selective forces that each act against rare alleles—interact paradoxically to increase the fixation probability of new mutant alleles. At sufficiently large, finite n, the fixation probability of a new underdominant allele under n-choice mating exceeds even that under random mating. For example, when selection is strong as $Ns = 10$, n-choice mating is beneficial to the rare underdominant allele provided the number of mating opportunities is a substantial fraction of the population, e.g. $n = 0.2N$. The stronger viability selection acts against heterozygotes, the smaller n is required for mate choice to facilitate fixation, relative to random mating.

This counterintuitive interaction between viability and sexual selection occurs because the manifold of equilibrium heterozygosity is reduced to $\tilde{h}(p) \approx 0$, for n large. Thus, there are few or no heterozygotes in the population on which viability selection can act. When there are many mating opportunities, a heterozygote parent is likely to find another heterozygote to mate with, and through Mendelian segregation this mating results in half the frequency of heterozygous progeny in each successive generation: roughly half of the minor alleles
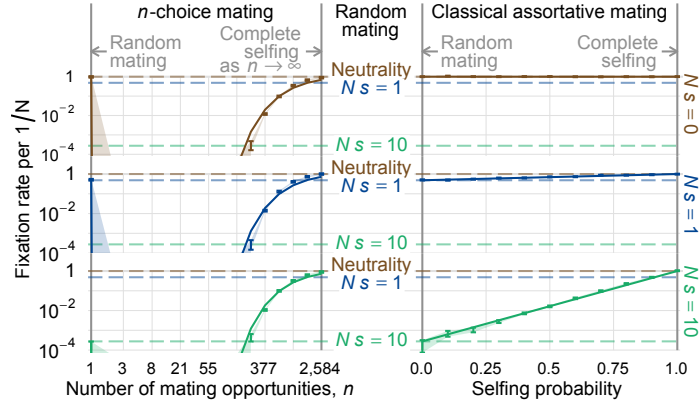
Figure 6: The effect of assortative mating on the fate of an underdominant allele, in the n-choice model (left) compared to the classical model of assortative mating by partial self-fertilization (right). The plot shows how the fixation probability of a new mutation present in one initial heterozygote depends on the number of mating opportunities, in the case of n-choice mating, or on the self-fertilization probability, in classical assortative mating. Vertical bars indicate the 95% confidence interval on the mean fixation rate observed in 100,000,000 simulated populations of size $N = Nf = 1,000$ under no viability selection ($Ns = 0$), weak viability selection ($Ns = 1$), and strong viability selection ($Ns = 10$). Dashed horizontal lines indicate the corresponding fixation probabilities of the underdominant allele under random mating. Fixation probabilities under the two models of assortative mating are equal either when $n = 1$ and the selfing probability is zero, or when $n \to \infty$ and the selfing probability equals one. Under classical assortative mating the fixation probability interpolates smoothly between these two limits. However, the n-choice model has a complex effect on fixation probabilities between these two limiting cases, and the probability may be higher or lower than random mating depending on n. Thus n-choice assortative mating can either impede or facilitate fixation of an underdominant allele.

present in heterozygotes are transferred to homozygotes at each mating. Thus with a sufficient number of mating opportunities, the mating system effectively hides hybrids from the eyes of underdominant selection.

1.5. Discussion

We have studied a mechanistic, one-parameter model of assortative mating that naturally spans the two classical extremes of random mating and complete assortment. The n-choice model fulfills the realistic condition that individuals can survey only a limited number of prospective mates. This simple formulation of mate selection nonetheless induces a complex form of sexual selection against rare alleles. In some regimes the induced selection is strong

enough to virtually prevent the fixation of rare alleles. If the locus guiding mate choice is also subject to underdominant viability selection then, provided the number of mating opportunities is large, the n-choice mating system can mask the effects of viability selection, greatly elevating the fixation rate of underdominant alleles in comparison to random mating.

The n-choice mating model provides a qualitatively different resolution to the puzzle of the observed fixation of underdominant alleles between populations. In a well-mixed population the fixation rate of an underdominant mutation decreases rapidly with population size. Wright's island model resolves this puzzle by exogenously subdividing the population into demes, so that fixation depends on the size of the demes rather than the whole population. For underdominant loci that also influence mate choice, the n-choice model can effectively decouple the fixation rate from population size without imposing a fixed population structure. Although both classical models of structured populations or partial selfing and the n-choice model facilitate fixation of underdominant alleles by suppressing heterozygosity, the mechanism and consequences of mating structure differ. In structured populations or partial selfing the inbreeding coefficient is exogenously fixed and it does not depend on allele frequency, whereas in the n-choice model the mating structure depends on the frequency of the rare allele so that, in particular, mating still occurs at random in monomorphic populations.

The precipitous drop in fixation probability of a novel mutant between random mating and n-choice mating, from $n = 1$ to $n = 2$ mating opportunities, is surprising. From the gestating parent's perspective, mate choice can only help rare alleles, as carriers of a rare allele copy have a greater chance of finding their own genotype to mate with and thus a lower chance of heterozygous offspring. However this gain of female function comes at a cost to male function. Because common types nearly always find their mates, but rare genotypes are more likely to settle for a common type, rare males are selected against. The relative strengths of reduced male fitness, heterozygote viability selection, and increased female fitness shift as the number of mating opportunities n increases, making preferred mates

more easily accessible. Female fitness is enhanced through reduction of viability selection against their offspring, and male fitness is less affected with higher values of n. The exact crossover point where fixation becomes more likely under n-choice mating than random mating depends on the strength of viability selection against heterozygotes, Ns, and the rate of participation in the mating system, Nf. Fixation probabilities that depend non-monotonically on a physical parameter are unusual, but they have also arisen, for different reasons, in models of subdivided populations with extinction and recolonization (Michalakis and Olivieri, 1993; Roze and Rousset, 2003).

The n-choice model can be seen as a (degenerate) case of the 'best-of-n' model introduced to study the efficiency of mate choice mechanisms (Janetos, 1980). Best-of-n models have been used in deterministic and stochastic settings to study the maintenance and efficiency of sexual selection (Seger, 1985; Pomiankowski, 1987) and speciation (Higashi et al., 1999; Arnegard and Kondrashov, 2004; Servedio and Bürger, 2014). Incorporating best-of-n mating into population-genetic models is known to produce different conclusions than under fixed relative preference assortative mating (Seger, 1985; Kuijper et al., 2012) for the efficiency of speciation (Arnegard and Kondrashov, 2004; M'Gonigle and FitzJohn, 2010; Servedio and Bürger, 2014). Despite these findings, most work on the fate of alleles in finite populations neglects mechanisms of non-random mate choice. The methods we have used to study the effects of n-choice mating in finite populations may be extended to other frequency-dependent mechanisms of mate choice and to other forms of viability selection and dominance on alleles.

We have described the n-choice model in the context of diploid parents, but minor variants of the model show similar behavior. For example, an analogous model in which a diploid parent censuses haploid microgametes, as occurs in flowering plants, has a different fixation probability in the limit $n \to \infty$ (see Fig. 7), but the same qualitative behavior remains: n-choice mating still induces sexual selection against rare types that can interact with underdominant viability selection to either impede or facilitate fixation of new mutants.
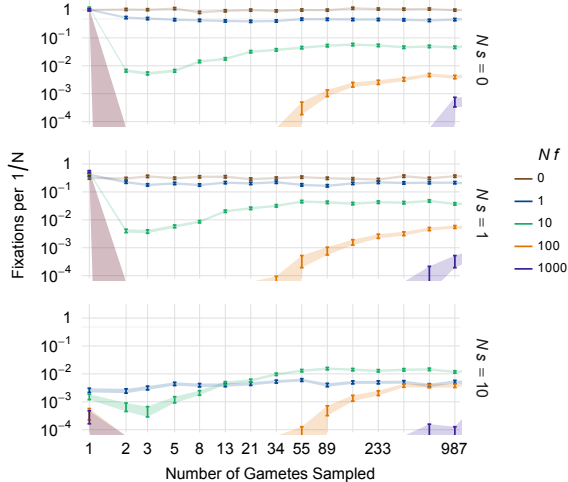
Figure 7: The fixation probability of one initial heterozygote in a model where females sample microgametes (sperm) attempting to raise homozygous offspring. This model is analogous to the n-choice model we study, but males are replaced by haplotypes. Females sample a limited number of gametes (n) and choose the first one that allows them to produce a homozygote, or, failing that, produce the heterozygote. The plot shows the fixation probability at different levels of viability selection against heterozygotes (Ns), and different rates of female participation in the mating system (Nf) on a background of clonal reproduction. Bands and error bars indicate the 95% confidence interval on the mean fixation rate in simulations with up to 100,000,000 runs in populations of size $N = 1,000$. For $Ns = 10$ and $Nf = 0$ (brown) the probabilities are below the range depicted. When females sample only one gamete ($n = 1$), the fixation probability is still roughly approximated by Eq. 1.1. At intermediate n, participation in the mating system induces strong fecundity selection against rare alleles. At large n, the fixation probability does not approach the neutral rate, because in order to form an initial mutant homozygote, an initial heterozygote must be chosen to reproduce, and it must also chose a mutant sperm instead of the more abundant wild-type. This depresses the fixation probability in the high-n limit relative to neutrality. Nonetheless, at large n and Ns the mating system can facilitate underdominant fixation.

Alternatively, if we prohibit selfing in the diploid model, n-choice mating is still more effective than random mating at fixing underdominant alleles at high n and Ns (Fig. 8). For simplicity, we have dealt only with symmetric underdominance, that is $w_{aa} = w_{AA}$. When this assumption is relaxed and one homozygote is preferable to the other, underdominance becomes a valley-crossing problem. Monte Carlo simulations exhibit qualitatively similar results here as well, with intermediate values of n retarding fixation of new mutations that are advantageous when homozygous but deleterious in heterozygotes, then facilitating fixation at sufficiently high n.
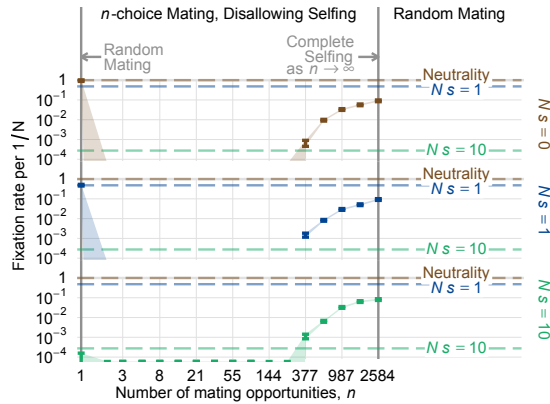
Figure 8: The effect of n-choice assortative mating on the fixation probability of an underdominant allele when self-mating is disallowed. The model and parameters are the same as those used in Figure 6, except that zygotes are drawn from a modified version of Eq. 1.2 which accounts for prohibition on self-mating. Vertical bars indicate the 95% confidence interval on the mean fixation rate observed in 100,000,000 replicate simulated populations of size $N = Nf = 1,000$ under no viability selection ($Ns = 0$, brown), weak viability selection ($Ns = 1$, blue), and strong viability selection ($Ns = 10$, green). Dashed horizontal lines indicate the corresponding fixation probabilities of the underdominant allele under random mating. The asymptotic fixation probabilities at high n are depressed relative to neutrality because an initial homozygote must first drift to copy number higher than 1 before its own genotype is available for mating.

The complex sexual selection induced by the n-choice mating model and its counterintuitive interaction with underdominant viability selection remind us that relaxing population-genetic assumptions can radically alter allele frequency dynamics in surprising ways. The astounding diversity of life-histories across taxa provides ample motivation to rethink conclusions drawn from standard models of randomly-mating diploid populations.

CHAPTER 2 : Drift and Selection in Culture

Text from this chapter has been published in *Nature* (Newberry et al., 2017).

Language and genes both evolve by transmission over generations with opportunity for differential replication of forms (Cavalli-Sforza and Feldman, 1981). The understanding that gene frequencies change at random by genetic drift, even without natural selection, was a seminal advance in evolutionary biology (Crow et al., 1970). Likewise, stochastic drift must also occur in language, due to randomness in how linguistic forms are copied between speakers (Bentley et al., 2004; Reali and Griffiths, 2010). Here we quantify the strength of selection relative to stochastic drift in language evolution. We use timeseries derived from large corpora of annotated texts spanning the 12th to 21st centuries to analyze many examples of three well-known grammatical changes in English: the regularization of past-tense verbs (Pinker, 1991; Lieberman et al., 2007; Michel et al., 2011; Reali and Griffiths, 2009; Hooper, 1976), the introduction of periphrastic 'do' (Ellegård, 1953), and variation in verbal negation (Jespersen, 1917). We reject stochastic drift in favor of selection in some cases but not others. In particular, we infer selection towards the irregular forms of some past-tense verbs, likely driven by changing frequencies of rhyming patterns over time. We show that stochastic drift is stronger for rare words, which explains why rare forms are more prone to replacement than common ones (Hooper, 1976; Lieberman et al., 2007; Pagel et al., 2007). This work provides a method to test selective theories of language change against a null model and reveals an underappreciated role for stochasticity in language evolution.

There is a rich history of exchange between linguistics and evolutionary biology, with August Schleicher among the earliest defenders of Darwinism (Schleicher, 1869; Darwin, 1888; Haeckel, 1868). Linguists have uncovered striking regularities in language change by examining which new forms enter a language and which forms are lost (Schleicher, 1869; Jespersen, 1917; Hooper, 1976; Bybee and Moder, 1983). Massive digital corpora (Michel et al., 2011; Davies, 2012) now provide precise frequency timeseries as one form replaces

another, which enable us to quantify evolutionary forces in language change using methods drawn from population genetics.

Language change involves competition between alternative linguistic forms, such as "sneaked" versus "snuck", whether they differ by sound, morphology, or syntactic structure (Cavalli-Sforza and Feldman, 1981; Lieberman et al., 2007; Pagel et al., 2007; Labov, 2001; Kroch, 1989; Christiansen et al., 2016). With every utterance, a speaker either invents a new form or uses one copied from other speakers. Forces that bias a speaker to adopt one form over another have been documented in rich detail (Croft, 2000). Examples include phonological analogy (Hooper, 1976; Prasada and Pinker, 1993), over-emphasis (Jespersen, 1917; Dahl, 2001), and a host of other social and cognitive factors (Labov, 2010; Hawkins, 1990). Any such bias in copying constitutes a form of selection in language evolution (Darwin, 1888). Explanations for language change, when one linguistic form increases in frequency and eventually replaces an alternative form over the course of generations, uniformly invoke selective mechanisms (Blythe and Croft, 2012; Kroch, 1989). However, the frequencies of alternative variants can change dramatically even without any bias towards one form or another, due to stochastic drift: randomness in the set of forms that each speaker happens to encounter and reproduce (Fig. 9). To determine the importance of directional forces we must first assess whether an observed language change is consistent with stochasticity in propagation alone.

Drift is recognized as an important null hypothesis in population genetics (Wright, 1931) and cultural evolution (Cavalli-Sforza and Feldman, 1981; Kandler and Shennan, 2013). More recently, linguists have suggested the use of null models for language change. Several models have been proposed (Bentley et al., 2004), including neutral evolution (Hahn and Bentley, 2003; Blythe, 2012); and some changes, such as new dialect formation, have been attributed to stochastic drift (Baxter et al., 2009). Yet methods to analyze drift versus selection in the vast available linguistic data have not been developed.

Here we systematically quantify drift and selection within well-known grammatical changes
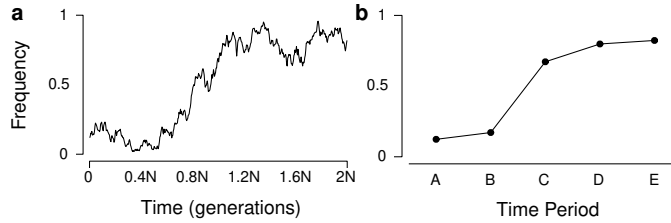
Figure 9: **A null model of language change.** Stochastic drift, random fluctuations in the frequencies of alternative forms, can accumulate to produce substantial change over time. We use the neutral Wright-Fisher diffusion from population genetics (Crow et al., 1970), which has also been derived as a model of language learning (Reali and Griffiths, 2010), as a null model of frequency variation due to stochastic drift. Panel **a** illustrates an example time-series of frequency variation produced by this null model. Although the complete time-series evidently shows random fluctuations, linguistic time-series are typically binned into time periods. When this time-series is binned (panel **b**), it produces a characteristic S-shaped curve that is often accepted as evidence of a directional force favoring one linguistic variant over others (Blythe and Croft, 2012; Kroch, 1989). This simple example illustrates the need to test hypotheses against a null model to definitively infer the presence of selective forces in language change (Blythe, 2012).

in English: the development of the morphological past tense in contemporary American English (Pinker, 1991; Cuskley et al., 2014) (spilt→spilled); the rise of the periphrastic 'do' in Early Modern English (Ellegård, 1953) (Mary ate not John's pizza → Mary did not eat John's pizza); and Jespersen's Cycle of sentential negation in Middle English (Jespersen, 1917) (Ic ne secge → I ne seye not → I say not). Our analyses are based on annotated texts ranging from the Norman conquest of England to the 21st century. In each case, we test whether observed linguistic changes are consistent with stochastic drift, or must involve selective forces. We compare the frequencies of alternative linguistic variants over time to predictions under the Wright-Fisher model of neutral stochastic drift. The Wright-Fisher model was first introduced in population genetics (Crow et al., 1970), but it has also been derived as a null model of linguistic change under Bayesian learning (Reali and Griffiths, 2010), where the inverse of the population size parameter N governs the amount of stochasticity in transmission, that is, the strength of drift ($1/N$).

We analyzed the evolution of past-tense verb conjugation by collecting verb tokens from the Corpus of Historical American English (Davies, 2012), which comprises more than four

Table 1: We analyzed 36 verbs with multiple past-tense forms appearing in the Corpus of Historical American English (Davies, 2012). The table shows each lemma, its corresponding regular and irregular forms, the number of times it occurs in the simple past tense in the corpus, and the FIT p-value for rejecting the neutral null hypothesis. The last three columns show the population size N and selection coefficient s of the regular form inferred by maximum likelihood in the two-parameter model (letting N and s vary), and the inferred population size N in the one-parameter model in which s is set to zero. A positive s indicates selection for the regular form (regularization), whereas negative s indicates selection against the regular form (irregularization) with strength given by the magnitude of s.

Lemma	Regular	Irregular	Count	FIT p-value	Inferred N	Inferred s	Inferred $N_{s=0}$
light	lighted	lit	8,869	0.003	1,247	-0.024	770
wake	waked	woke	7,186	0.006	1,714	-0.024	4,005
sneak	sneaked	snuck	898	0.02	4,135	-0.039	29
weave	weaved	wove	907	0.02	1,052	0.013	126
dive	dived	dove	1,036	0.05	2,148	-0.018	710
smell	smelled	smelt	4,555	0.05	974	0.038	1,708
wed	wedded	wed	211	0.08	888	-0.026	154
quit	quitted	quit	2,734	0.10	346	-0.048	430
spell	spelled	spelt	962	0.10	1,556	0.025	2,055
shine	shined	shone	8,424	0.10	20,110	0.019	989
tell	telled	told	129,041	0.12	20,930	-0.050	314,100
leap	leaped	leapt	8,336	0.13	2,003	-0.019	346
build	builded	built	9,109	0.14	87	-0.063	8,602
plead	pleaded	pled	3,810	0.14	8,050	-0.006	3,756
learn	learned	learnt	18,851	0.16	2,467	0.027	6,918
heave	heaved	hove	2,392	0.20	2,371	0.008	3,663
dwell	dwelled	dwelt	1,621	0.20	16,170	0.033	63
wet	wetted	wet	770	0.24	5,707	-0.009	7,903
burn	burned	burnt	6,097	0.24	7,213	0.014	9,045
kneel	kneeled	knelt	5,185	0.30	7,299	-0.006	8,912
spoil	spoiled	spoilt	1,045	0.30	3,519	0.018	3,431
awake	awaked	awoke	4,926	0.32	1,676	-0.036	15,860
know	knowed	knew	171,518	0.34	9,397	-0.007	18,980
speed	speeded	sped	3,142	0.39	867	-0.002	1,077
lay	laid	lay	66,436	0.45	10,610	-0.002	11,070
spill	spilled	spilt	1,178	0.47	1,266	0.031	1,509
throw	throwed	threw	25,612	0.61	729	-0.011	2,001
strew	strewed	strew	727	0.66	1,539	0.000	1,537
hear	heared	heard	72,052	0.72	1,129	-0.033	28,530
knit	knitted	knit	675	0.76	173	-0.010	333
hang	hanged	hung	16,079	0.77	3,855	-0.012	12,450
dream	dreamed	dreamt	4,530	0.77	2,832	0.005	3,907
catch	catched	caught	25,529	0.78	22,520	-0.021	83,060
lean	leaned	leant	16,981	0.85	2,166	0.013	7,594
draw	drawed	drew	35,213	0.87	2,102	-0.026	19,620
grow	growed	grew	37,444	0.93	7,094	-0.013	21,340
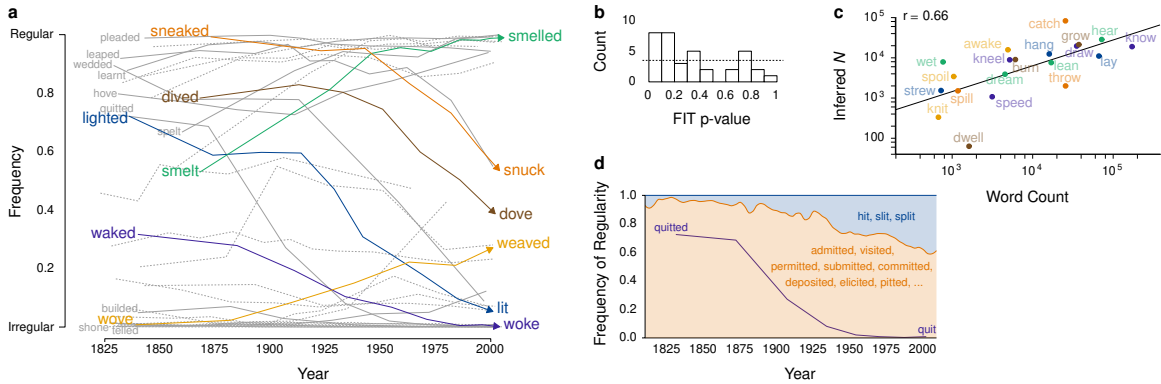
Figure 10: **Verb regularization and irregularization.** We analyzed 36 verbs with multiple past-tense forms appearing in the Corpus of Historical American English (Davies, 2012). Six of these verbs (**a**, colored lines) experience selection for either regularization or irregularization, each with nominal $p<0.05$ by the Frequency Increment Test of selection (FDR=30%). The regular form is favored in two of these cases, and the irregular form in the remaining four cases. Ten more verbs (**a**, solid gray lines), of which four are regularizing, are significant at specificity $1 - \alpha = 0.8$, with a false discovery rate of 45%. The distribution of nominal FIT p-values (**b**) is non-uniform (Kolmogorov-Smirnov $p=0.002$), which confirms that some verbs experience selection. Changing use of rhyming patterns may drive selection for irregular forms, such as quitted $\rightarrow$ quit (**d**), whose irregularization coincides with increasing use of the irregular verbs hit, slit, and split (see also Fig. 11). Among the remaining 20 verbs most consistent with neutrality (**a**, dashed gray lines), the log inferred population size assuming $s=0$ correlates with log token count in the corpus (**c**, Pearson $r=0.66$, $p=0.002$).

hundred million words from $>$100,000 texts between the years 1810 and 2009, tagged for part of speech. Among all tokens assigned the simple past tense as the most likely part of speech, we selected those lemmas with two past-tense variants each occurring at least 50 times in the corpus (SI Text). This produced 704,081 tokens that provide frequency timeseries for the regular versus irregular forms of 36 polymorphic verbs (Fig. 10) which range from very rare ("wed", one in two million) to common ("know", one in two thousand). For each timeseries we computed a two-sided p-value for rejecting neutral stochastic drift by the Frequency Increment Test (FIT, see SI Text). We also inferred the most likely population size, N, and selection coefficient, s, favoring one linguistic variant over another (SI Text, Table 1).

We conclude that selection is driving changes in past-tense conjugation among six of the

polymorphic verbs, each with nominal $p < 0.05$. In four of these cases selection favors the irregular variant (lighted→lit, waked→woke, sneaked→snuck, dived→dove), whereas two cases exhibit regularization (wove→weaved, smelt→smelled). To account for multiple hypothesis testing we computed the false-discovery rate among these six verbs: FDR=30%. Furthermore, we note that the distribution of all FIT p-values is non-uniform (Kolmogorov-Smirnov p=0.002, Fig. 10b), which confirms that selection is operating on some of the polymorphic verbs.

Selection for regularization comes as no surprise. Prevailing linguistic theory predicts regularization (Hooper, 1976; Pinker, 1991), for reasons of economy or cognitive ease (Jakobson et al., 1995; Zipf, 1949; Pinker, 1991). Trends towards past-tense regularization have been observed, especially for rare words over long timescales from Old to Modern English (Hooper, 1976; Michel et al., 2011; Lieberman et al., 2007). We identify cases of incipient regularization, such as wove→weaved, where the regular variant is in the minority at present but predicted by our analysis to eventually replace the irregular form.

Selection for irregularization is more mysterious, although several cases have been noted (Bybee and Moder, 1983; Michel et al., 2011; Cuskley et al., 2014). In Modern English, we find that irregularization is as common as regularization (Fig. 10). One possible explanation involves rhyming. Psychological studies have found speakers willing to copy or invent irregular variants, such as spling/splung (Prasada and Pinker, 1993), that rhyme with existing irregular verbs (Ullman, 1999). Our analysis of dive/dived versus dive/dove reveals selection for dive/dove, which coincides with a marked increase of the irregular verb drive/drove in the corpus, associated with the invention of cars in the 20th century. More generally, the inferred selection coefficient (s) always favors the irregular variant of a polymorphic past-tense verb when similar-sounding irregular verbs are on the rise in the corpus (11 out of 11 cases: light, dive, quit, tell, leap, build, kneel, know, throw, knit and grow; see Table 1, Fig. 11). For example, selection for quitted→quit coincides with increased use of the irregular verbs hit/hit and split/split (Fig. 10d). The frequency of "split" increased
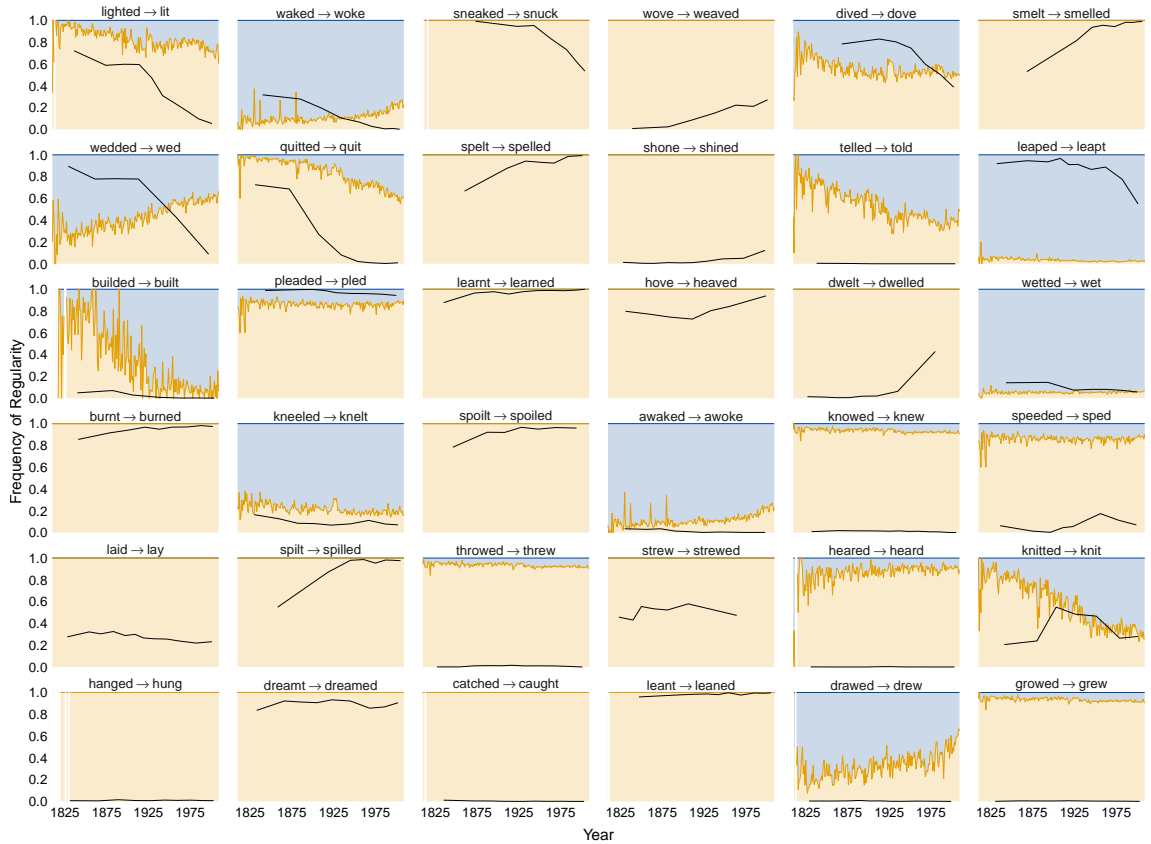
Figure 11: **Timeseries of changing rhyming patterns.** Each panel shows the time-series of a polymorphic verb (black lines), repeated from Fig. 10a, and the frequency of similar-sounding monomorphic regular (orange) and irregular (blue) verbs in the Corpus of Historical American English. The tokens included are all tenses of those lemmas with a pronunciation known to the CMU Pronouncing Dictionary in both the lemma and the simple past tense. The list of verbs incorporated in each timeseries is given in Table 3. For 17 polymorphic verbs we find no similar-sounding monomorphic irregular verbs (all-orange panels). The title of each panel indicates the sign of the maximum-likelihood selection coefficient, either regular→irregular or irregular→regular.

53

nearly four-fold over the past century as split acquired an additional meaning (*to leave or depart*). Thus a semantic change in one irregular verb ("split") likely induced selection for irregularization in another, semantically unrelated verb ("quit") that shares the same past-tense rhyming pattern.

Selection towards an irregular variant can also occur when similar-sounding irregular verbs are on the decline, as in the case of wedded→wed (Fig. 11). Our inference of selection for wedded→wed is particularly striking because it contradicts previous work predicting the regularization of wed→wedded based on long-term trends (Lieberman et al., 2007).

Drift alone is sufficient, however, to explain the observed changes for the majority of verbs we analyzed in Modern English (Table S1, FIT p values). These include verbs previously described as undergoing regularization, such as spilt→spilled and burnt→burned (Michel et al., 2011; Cuskley et al., 2014) (SI Text). Failure to reject neutrality in these cases does not imply that selection is entirely absent. For example, there is likely some selection for knitted→knit due to rhyming, as with quitted→quit (Fig. 11). Nonetheless, the inferred strength of selection for "knit" is too weak relative to drift to affect its dynamics: $|Ns|=1.67$ for "knit" (FIT $p=0.76$) in contrast to $|Ns|=30.51$ for "lit" (FIT $p=0.003$) (Table S1). Even with some amount of selection, dynamics are indistinguishable from neutrality when drift is strong enough (Crow et al., 1970).

Among the verbs whose dynamics are dominated by drift, the strength of drift correlates inversely with the verb's overall frequency in the corpus (Fig. 10). This result implies that common words should exhibit less variability over time than rare words—a phenomenon that has indeed been observed in many empirical studies (Hooper, 1976; Lieberman et al., 2007; Pagel et al., 2007) and previously attributed to stronger purifying selection against novel variants of common words (Pagel et al., 2007). Our analysis provides an alternative and complementary explanation for faster rates of replacement in rare words: rare words, whether under selection to change or not, experience more stochasticity in transmission. Furthermore, this explanation predicts that the replacement of one form by another in a

Table 2: We analyzed the rise of *do*-support in Early Modern English in four grammatical contexts using instances of potential *do*-support from the Penn Parsed Corpora of Historical English. The table indicates each context, the corresponding number of tokens of potential *do*-support in the corpus, the FIT *p*-value for rejecting the neutral null hypothesis, the inferred population size N, and the inferred selection coefficient s in favor of *do*-support.

Context	Count	FIT p-value	Inferred N	Inferred s
negative interrogative	606	0.270	504	0.013
affirmative interrogative	4,665	0.178	3,545	0.006
negative declarative	14,227	0.005	2,477	0.014
negative imperative	1,231	0.003	23,150	0.014

rare word is more likely to occur by random chance, whereas such substitutions in common words are more likely caused by selection.

Next we analyzed the rise of *do*-support in Early Modern English, as the auxiliary verb 'do' comes to express the tense of a sentence (Ellegård, 1953). Over the course of centuries, for example, "You said not." becomes "You did not say." and "Asked you a question?" becomes "Did you ask a question?". We collected instances of potential *do*-support from the Penn Parsed Corpora of Historical English (SI Text). This dataset includes roughly seven million syntactically parsed words from 1,220 texts of British English, offering a much larger timeseries than previous work (Ellegård, 1953). We extracted 20,729 instances of potential *do*-support in the context of affirmative questions, negative questions, negative declaratives, and negative imperatives.

We find that the rise of the periphrastic 'do' was more rapid in negative declarative and imperative statements, where we reject drift (p=0.005 and p=0.003 respectively), than in affirmative questions, where we fail to reject drift (p=0.18, Fig. 12). *do*-support also appears to rise suddenly in negative questions, although the force of drift is strongest in this case (Table 2), and indeed we fail to reject drift (p=0.27, Fig. 12). We might expect that selection for an auxiliary verb will operate equally in all grammatical contexts (Kroch, 1989), and yet the extensive parsed corpora available do not support this hypothesis. Our analysis suggests an alternative scenario: the periphrastic 'do' first drifted by chance to high frequency in questions, which then induced a directional bias towards 'do' in declarative and imperative
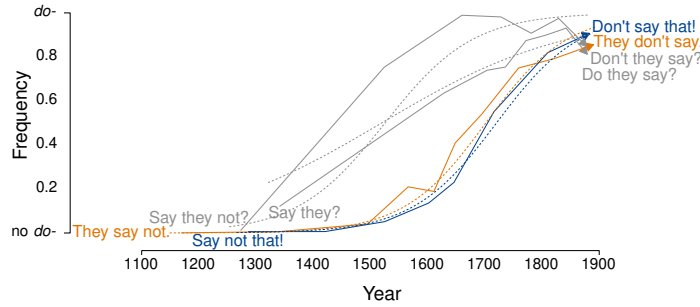
Figure 12: **The rise of the periphrastic 'do' in early Modern English.** The frequency of 'do' as an auxiliary verb first rose in the context of interrogative sentences (gray). However, we cannot reject drift for either affirmative interrogatives (FIT $p=0.18$) or negative interrogatives (FIT $p=0.27$). Subsequently, *do*-support rose rapidly in negative declarative and negative imperative sentences, where we detect selection (FIT $p=0.005$ and $p=0.003$, respectively). Dotted lines plot the logistic curve with slope determined by the maximum-likelihood selection coefficient inferred in each grammatical context (Table 2). These results suggest *do*-support rose by chance through drift in interrogative statements, setting the stage for directional evolution of *do*-support in other grammatical contexts.

statements, for reasons of grammatical consistency or cognitive ease.

Finally, we studied the evolution of syntactic verbal negation from the 12th to the 16th centuries, based on 5,475 negative declaratives extracted from the Penn Parsed Corpus of Middle English. We observe pre-verbal negation (e.g. "Ic ne secge") giving way to embracing bipartite negation ("I ne seye not") and then finally to post-verbal negation ("I say not"), in a pattern known as Jespersen's Cycle (Jespersen, 1917). For both transitions that form this cycle we definitively reject neutral drift (FIT $p<0.05$, Fig. 13). This provides statistical support for longstanding hypotheses that changes in verbal negation are driven by directional forces, such as phonetic weakening (Jespersen, 1917), or a tendency for speakers to over-use more emphatic forms of negation (Jespersen, 1917; Dahl, 2001) which then lose their emphasis as they become dominant (Crawford and Sobel, 1982; Dahl, 2001). Although directionality in Jespersen's cycle was first recognized by comparing multiple languages (Jespersen, 1917), we reach the same conclusion by analyzing changes in English alone.

Methods drawn from phylogenetics have enabled researchers to infer the relationships among divergent languages (Ringe et al., 2002; Gray and Atkinson, 2003; Pagel, 2013, 2009; Pagel
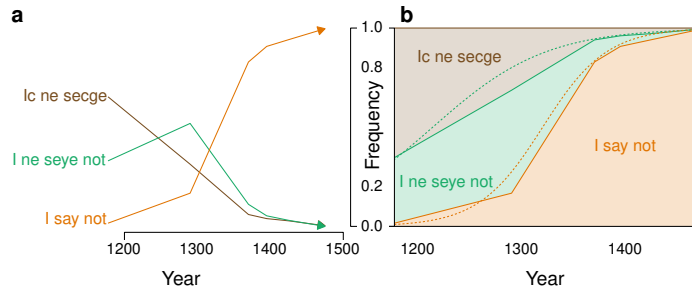
Figure 13: **Evolution of verbal negation.** In English and French, pre-verbal negation (e.g. Old English "Ic ne secge") gave way to embracing bipartite negation (Middle English "I ne seye not") and then to post-verbal negation (Early Modern English "I say not"), in a pattern known as Jespersen's Cycle. We plot the frequencies of these forms among 5,918 instances of negation from 56 texts in the Penn-Helsinki Parsed Corpus of Middle English (**a**). We infer selection for bipartite and post-verbal negation in the background of pre-verbal forms (FIT p=0.02, **b**, green lines) and selection for post-verbal negation in a mixed population of pre-verbal and bipartite forms (FIT p=0.04, **b**, orange lines). Dotted lines indicate logistic curves corresponding to maximum-likelihood selection coefficients.

et al., 2007). By contrast, the study of how a language changes over short timescales has not taken full advantage of statistical inference. And yet, changes within a language must be the origin of differentiation between languages (Lupyan and Dale, 2015). The combination of massive digital corpora along with timeseries inference techniques from population genetics now allows us to disentangle distinct forces that drive language evolution. How exactly individual-level cognitive processes in a language learner (Pinker, 1991; Jespersen, 1917; Kroch, 1989; Zipf, 1949; Jakobson et al., 1995) produce population-level phenomena such as drift and selection (Tamariz et al., 2014) remains a topic for future research.

Source code to reproduce our results from published databases is available online at `http://github.com/mnewberry/ldrift`.

2.1. Frequency Increment Test for selection.

To test for selection in an empirical timeseries, we apply a transformation that produces homoscedastic frequency increments under the null hypothesis of stochastic drift. The mean transformed frequency increment can then be compared to zero, its expectation under the null, using Student's t-test, which we call the Frequency Increment Test (FIT) (Feder et al.,

2014). This procedure tests the composite null hypothesis of stochastic drift even when the strength of drift is unknown ($s=0$, N arbitrary), against the alternative hypothesis that the mean frequency increment differs from zero. The alternative hypothesis makes no assumption about the form of selection, be it frequency-dependent, time-dependent, or correlated with other variables or between forms. Directional change induced by repeated innovation (biased mutation) or immigration also would cause the mean frequency increment to differ from zero. Hence the FIT tests a null hypothesis of neutral drift against an alternative hypothesis of some directional force influencing the course of evolution. This test is valid for a large class of neutral null models: all those with the same diffusion limit as the Wright-Fisher model (Ewens, 2004). Many models of language change also share this diffusion limit (Reali and Griffiths, 2010; Baxter et al., 2006; Fontanari and Perlovsky, 2004), whether variation arises during first language acquisition or in usage over the lifetime of an individual speaker, although spatial models (Nettle, 1999) do not.

The Wright-Fisher diffusion is a Markov process with independent increments that are normally distributed over infinitesimal increments in time. A Gaussian approximation to the diffusion process is valid away from the boundary frequencies zero and one (Ewens, 2004). Under the Gaussian approximation to the neutral diffusion, the probability $\mathbb{P}(x_{i+1}, t_{i+1}|x_i, t_i)$ of observing an allele at frequency x_{i+1} at time t_{i+1} given frequency x_i at time $t_i < t_{i+1}$ follows the normal distribution:

$$\mathbb{P}(x_{i+1}, t_{i+1}|x_i, t_i) \sim \text{Normal}(x_i, 2x_i(1 - x_i)(t_{i+1} - t_i)/N). \tag{2.1}$$

The Frequency Increment Test compares the distribution of observed frequency changes in an empirical timeseries with this null distribution.

Assume we have a series of frequencies, $x_1, \ldots, x_n$ of one linguistic variant relative to another, observed at n times ordered $t_1, \ldots, t_n$. The FIT compares the frequency increments between subsequent pairs of observations to the null distribution specified by the Gaussian approximation to the neutral diffusion. To make this comparison, we first rescale each

frequency increment $x_{i+1} - x_i$. The rescaled increments Y_i are given by

$$Y_i = \frac{x_{i+1} - x_i}{\sqrt{2x_i(1 - x_i)(t_{i+1} - t_i)}},\tag{2.2}$$

and thus the Y_i are independent with $Y_i \sim \text{Normal}(0, 1/N)$ under the null. Recall that, under the composite null hypothesis of neutral drift the population size is unspecified, so that $1/N$ must be treated as a nuisance parameter. The classical t-test solves this problem with maximal power. To perform the FIT we simply calculate the t-statistic for the rescaled frequency increments $Y_1, \ldots, Y_{n-1}$. We reject neutrality for any size population if the quantile of this absolute t-statistic exceeds $1 - \alpha/2$ under the t distribution with $n - 2$ degrees of freedom.

The test of selection described above controls the type I error rate but not the type II error rate. Failure to reject drift even when selection is in fact present can occur either because of insufficient sample size and power (although we can rule this out in several cases, see below), or because the strength of selection (s) is small compared to that of drift $(1/N)$ so that drift dominates the dynamics. Drift also dominates selection when the frequencies are near the boundary zero or one (Ewens, 2004). Finally, FIT tests for directional selection, and it may produce an insignificant result when forms experience balancing selection for a stable intermediate frequency.

2.2. Estimating population size and selection coefficients.

Given a series of frequencies $x_1, \ldots, x_n$ observed at times $t_1, \ldots, t_n$ we use maximum likelihood to estimate the population size, N, and the selection coefficient, s, favoring one linguistic form over another. In the presence of selection, this computation involves approximating the diffusion by a deterministic path plus Gaussian noise (Pollett, 1990). The deterministic expectation follows a logistic curve that depends upon the choice of s, whereas the variance of the Gaussian noise depends on both s and N. We optimize the likelihood of the data under this model, varying N and s, using `tsinfer` software by Feder et al. (2014)

59

(`https://github.com/skryazhi/tsinfer`), which identifies the most likely parameters by the Nelder-Mead simplex optimization method, treating the sampled frequencies as true frequencies (`-F` flag to `tsinfer`).

The most likely population size estimated by this procedure is known only up to a fixed, constant factor, the generation time (Ewens, 2004). We chose this unit of time to be 1 year for purposes of presentation.

Note that the maximum-likelihood estimate of s will almost always be non-zero. Confidence intervals on s may exclude zero even when FIT fails to reject neutrality. However, confidence intervals on s inferred from the likelihood surface by the likelihood-ratio statistic are known to be biased (Feder et al., 2014). We therefore use the conservative FIT to test for selection against the composite null hypothesis of neutrality.

2.3. Analysis of past-tense verbs.

We collected all tokens from the Corpus of Historical American English (Davies, 2012) (COHA) tagged past-tense verb (`vvd`) as most likely part of speech, retaining those with two past-tense variants each occurring at least 50 times in the corpus. We excluded verbs with temporal variation caused by spelling conventions (e.g cancelled versus canceled), lemmas with semantic ambiguities (e.g. bear versus bore, wind), and lemmas with multiple irregular variants (e.g. begin, bid, drink, ring).

For each verb, we assigned time points and frequencies by binning its tokens into date ranges of variable lengths to ensure roughly the same number of tokens per bin (quantile binning). We set the number of bins equal to the logarithm of the number of tokens, rounded up, and we assigned to each bin the midpoint of its date range as its time point. The observed frequency of the regular form is taken as the proportion of the regular form in each bin, and the time is the midpoint of the bin. We applied Laplace (add-one) smoothing to counts with only one variant present in a bin, in order to remove apparent absorption events (Feder et al., 2014). We tested for the presence of selection in each verb using the FIT (see Table 1). We

tested for the presence of selection in at least some of the verbs with a Kolmogorov-Smirnov test for uniformity of the empirical distribution of FIT p-values (see Table 1, Fig. 10b).

Results of the Frequency Increment Test for selection on these 36 verbs are not driven by power or sample size. There is no significant difference in the mean number of tokens among the 16 verbs with FIT $p<0.2$ compared to the remaining verbs (Mann-Whitney test, $p=0.29$). Conversely, the failures to detect selection highlighted in the main text are not due to small sample sizes: the number of tokens for spill (1,178) exceeds that of four verbs with FIT $p<0.2$; and the number of tokens for burn (6,097) exceeds that of eight verbs with FIT $p<0.2$ (see Table 1).

We also controlled for fluctuations in the total usage of each verb over time (Fig. 14). We view the regular and irregular simple past-tense variants of a polymorphic verb as competing forms, in the sense that only one or the other can be used in any past-tense utterance. The total use of each verb in the past tense nonetheless fluctuates over time, as measured by its frequency among all tokens in the corpus. In fact, the total use of each verb in any tense also fluctuates over time (Fig. 14), and the fluctuations in total use are highly correlated with fluctuation in past-tense use ($r^2=0.992$). In other words, even though past-tense usage may vary over time, its variation simply reflects variation in the overall usage of the verb. We find no significant correlation between the strength of the trend in past-tense usage (log-linear regression slope) with the inferred strength of selection, $|s|$, on the regular form ($r^2=0.05$, $p=0.18$). Fluctuations in the total use of a verb over time are thus analogous to fluctuating population size, and our inferred N measures the effective population size (Ewens, 2004).

2.4. Analysis of rhyming patterns.

To examine one possible source of selection towards irregularization we extracted verbs that share the same rhyming pattern and tracked the token frequencies of regular and irregular rhyming patterns in the corpus over time. For each of the 36 polymorphic verbs,
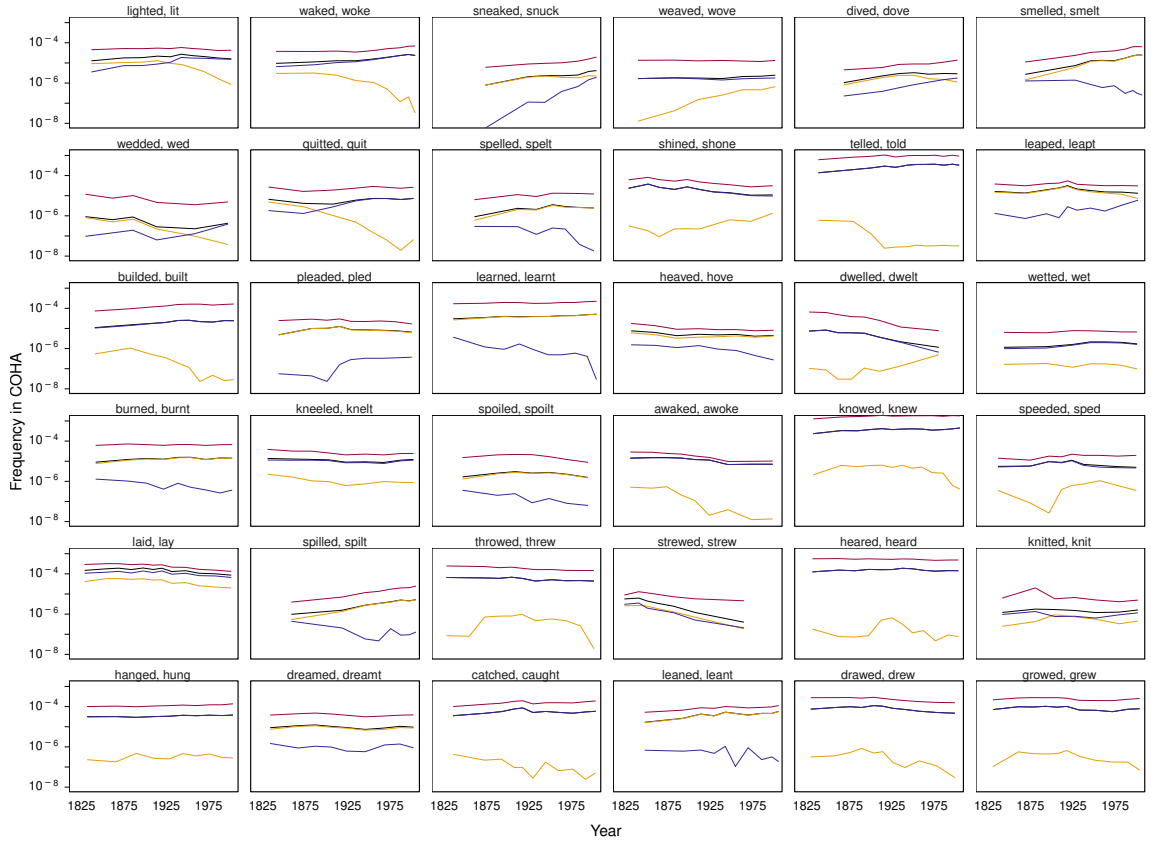
Figure 14: **Temporal trends in the usage of 36 verbs, in the simple past tense and in all tenses.** The figure shows the frequency of each verb in all tenses (red), in the simple past tense (black), and in the regular (blue) and irregular (yellow) variants of the simple past. For each verb, the trend in usage in the past tense (black) is virtually the same as the trend in all tenses (red), with the notable exception of "strew". As a result, we do not expect that any trend in overall past-tense usage should correlate with differential use of the regular versus irregular past-tense variant. Indeed, we find no significant correlation between the strength of the trend in past-tense usage (log-linear regression slope) with the inferred strength of selection, $|s|$, on the regular form (r^2=0.05, p=0.18).

Table 3: **List of similar-sounding monomorphic verbs for each past-tense conjugation of polymorphic verbs.** For each polymorphic verb we study, its regular and irregular variants each define a present/past rhyme scheme. A monomorphic lemma is considered similar-sounding if it rhymes in both its lemma and its past tense with a given rhyme scheme. A lemma is included in the table if 1) it has only one simple past-tense form occurring over 50 times in the Corpus of Historical American English, 2) both the past-tense form and the lemma itself are present in the CMU Pronouncing Dictionary, and 3) the lemma and past-tense forms fit either the regular or irregular rhyme scheme of one of the polymorphic verbs.

| Lemma | Rhyme scheme | | Similar-sounding lemmas | |
	Regular	Irregular	Regular	Irregular
light	light/lighted	light/lit	invite, excite, delight, cite	bite
wake	wake/waked	wake/woke	stake, rake, fake, brake, bake, ache	break
sneak	sneak/sneaked	sneak/snuck	wreak, tweak, streak, squeak, shriek, pique, peek, peak, leak, freak, creak	
weave	weave/weaved	weave/wove	retrieve, relieve, receive, perceive, grieve, deceive, conceive, believe, achieve	
dive	dive/dived	dive/dove	thrive, survive, revive, derive, deprive, contrive, arrive	strive, drive
smell	smell/smelled	smell/smelt	yell, swell, shell, repel, rebel, quell, propel, parallel, impel, fell, expel, excel, dispel, compel	
wed	wed/wedded	wed/wed	thread, spearhead, shred, imbed, head, dread, behead	spread, shed
quit	quit/quitted	quit/quit	visit, submit, spirit, solicit, remit, pit, permit, outwit, outfit, forfeit, elicit, deposit, credit, counterfeit, commit, benefit, admit, acquit	split, slit, hit
spell	smell/smelled	smell/smelt	yell, swell, shell, repel, rebel, quell, propel, parallel, impel, fell, expel, excel, dispel, compel	
shine	shine/shined	shine/shone	undermine, underline, twine, sign, resign, refine, redesign, redefine, recline, pine, outline, opine, mine, malign, line, incline, headline, fine, entwine, dine, design, define, decline, consign, confine, combine, assign, align	
tell	smell/smelled	tell/told	yell, swell, shell, repel, rebel, quell, propel, parallel, impel, fell, expel, excel, dispel, compel	sell
leap	leap/leaped	leap/leapt	seep, reap, peep, heap, beep	weep, sweep, sleep, keep, creep
build	build/builded	build/built	gild	rebuild
plead	plead/pleaded	plead/pled	supersede, succeed, stampede, seed, secede, recede, proceed, precede, need, intercede, impede, heed, exceed, concede, cede, bead, accede	mislead, feed, breed, bleed
learn	learn/learned	learn/learnt	yearn, turn, sunburn, spurn, return, overturn, iron, govern, earn, discern, concern, churn, adjourn	
heave	weave/weaved	weave/wove	retrieve, relieve, receive, perceive, grieve, deceive, conceive, believe, achieve	
dwell	smell/smelled	smell/smelt	yell, swell, shell, repel, rebel, quell, propel, parallel, impel, fell, expel, excel, dispel, compel	
wet	wet/wetted	wet/wet	whet, sweat, silhouette, regret, pet, net, jet, fret, abet	upset, set, offset, let, bet, beset
burn	learn/learned	learn/learnt	yearn, turn, sunburn, spurn, return, overturn, iron, govern, earn, discern, concern, churn, adjourn	
kneel	kneel/kneeled	kneel/knelt	wheel, unseal, steel, seal, reveal, repeal, reel, peel, heel, heal, congeal, conceal, appeal	feel, deal
spoil	spoil/spoiled	spoil/spoilt	toil, soil, roil, recoil, oil, foil, coil, broil, boil	

	Rhyme scheme		Similar-sounding lemmas	
Lemma	Regular	Irregular	Regular	Irregular
awake	wake/waked	wake/woke	stake, rake, fake, brake, bake, ache	break
know	know/knowed	know/knew	zero, winnow, wallow, veto, tow, torpedo, toe, swallow, stow, sow, snow, slow, show, shadow, sew, row, radio, owe, overshadow, overflow, narrow, mow, hollow, glow, furrow, foreshadow, follow, flow, elbow, echo, crow, bow, borrow, billow, bestow, bellow	overthrow, blow
speed	plead/pleaded	plead/pled	supersede, succeed, stampede, seed, secede, recede, proceed, precede, need, intercede, impede, heed, exceed, concede, cede, bead, accede	mislead, feed, breed, bleed
lay	lay/laid	lay/lay	weigh, waylay, underpay, sway, survey, stray, stay, spray, ricochet, replay, repay, relay, pray, portray, play, pay, parlay, outweigh, obey, flay, display, disobey, dismay, delay, decay, convey, buffet, betray, bay, array, allay	
spill	spill/spilled	spill/spilt	will, thrill, still, stencil, refill, mill, kill, grill, fulfil, fill, drill, distil	
throw	know/knowed	know/knew	zero, winnow, wallow, veto, tow, torpedo, toe, swallow, stow, sow, snow, slow, show, shadow, sew, row, radio, owe, overshadow, overflow, narrow, mow, hollow, glow, furrow, foreshadow, follow, flow, elbow, echo, crow, bow, borrow, billow, bestow, bellow	overthrow, blow
strew	strew/strewed	strew/strew	woo, view, value, undervalue, tattoo, sue, subdue, stew, spew, shoo, screw, review, rescue, renew, pursue, issue, interview, glue, eschew, ensue, discontinue, debut, coo, continue, construe, chew, brew, boo, argue, accrue	
hear	hear/heared	hear/heard	disappear	overhear
knit	knit/knitted	quit/quit	transmit, omit, fit, emit, budget	split, slit, hit
hang	hang/hanged	hang/hung	harangue, bang	
dream	dream/dreamed	dream/dreamt	team, stream, steam, seem, scream, scheme, redeem, gleam, esteem, deem, cream, beam	
catch	catch/catched	catch/caught	snatch, scratch, patch, match, latch, hatch, detach, attach	
lean	lean/leaned	lean/leant	wean, screen, preen, intervene, glean, convene, clean, careen	
draw	draw/drawed	draw/drew	thaw, saw, paw, outlaw, gnaw, claw, awe	withdraw
grow	know/knowed	know/knew	zero, winnow, wallow, veto, tow, torpedo, toe, swallow, stow, sow, snow, slow, show, shadow, sew, row, radio, owe, overshadow, overflow, narrow, mow, hollow, glow, furrow, foreshadow, follow, flow, elbow, echo, crow, bow, borrow, billow, bestow, bellow	overthrow, blow

we extracted all monomorphic verbs that rhyme in both the lemma and the past tense with either the regular (e.g. dive/dived) or irregular (e.g. dive/dove) form of conjugation. We considered a verb to be monomorphic when only one past-tense form was present at least 50 times in COHA (e.g. thrive/thrived, or strive/strove). We used the CMU Pronouncing Dictionary to match rhyming words, ignoring stress. For each polymorphic verb, the set of monomorphic lemmas that rhyme with either its regular or irregular conjugation is listed

in Table 3.

For each pair of rhyming patterns associated with the regular and irregular forms of a polymorphic verb, we created an annual timeseries of the frequency of regular among regular plus irregular monomorphic lemmas in COHA (Fig. 11). We classified the regular rhyming pattern as decreasing in frequency if the slope of the timeseries (by linear regression) was negative. This slope was negative for 11 of the polymorphic verbs (light, dive, quit, tell, leap, build, kneel, know, throw, knit and grow) and positive for 8 polymorphic verbs (wake, wed, plead, wet, awake, speed, hear and draw). The remaining 17 polymorphic verbs had no similar-sounding irregular lemmas in the corpus, and so the slope was zero.

For all 11 polymorphic verbs whose regular rhyming pattern had decreasing frequency in the corpus, the maximum-likelihood selection coefficient was negative (Table 1), favoring the irregular form of the verb. Moreover, among all 36 polymorphic verbs, the slope of the rhyming pattern over time is correlated with the inferred selection coefficient (Pearson $r=0.35$, $p=0.04$).

Although we tracked the token frequency of lemmas that rhyme in both past and present form following Prasada and Pinker (1993), other research has suggested rhyming past-tense forms alone (Bybee and Moder, 1983), type frequency (Bybee, 1985), or a combination of type and token frequency (Hay and Baayen, 2002) would more closely approximate underlying psychological mechanisms of language learning and use. We did not use type frequency in our analysis, as the number of types is too few to register change over time. We can, however, relax our notion of similar rhyming patterns to include any verbs that rhyme in the past tense. In this case, the slope of rhyming pattern frequency over time is not significantly correlated with the inferred selection coefficient (Pearson $r=0.20$, $p=0.25$). In this larger set of past-tense verbs, it is possible to register some change in type frequencies, but this does not improve the correlation with inferred selection coefficients (Pearson $r=-0.02$, $p=0.92$).

2.5. Analysis of *do*-support.

We analyzed syntactically parsed clauses from the Penn-Helsinki Parsed Corpus of Middle English (1100-1500), the York-Helsinki Parsed Corpus of Early English Correspondence (1400-1700), the Penn-Helsinki Parsed Corpus of Early Modern English (1500-1700), and the Penn Parsed Corpus of Modern British English (1710-1910), excluding works of philosophy, translations of Boethius, sermons, biblical translations, and statutes as unrepresentative of the language of their times. Our timeseries is larger, covers more of the rise of *do*-support, and samples from substantially more authors per token than the data hand-collected by Ellegård (1953). We used the CorpusSearch 2 tool (version 2.003.04, Beth Randall, Ann Taylor and Anthony Kroch, `http://corpussearch.sourceforge.net/`, 2005-2013) to label all clauses for mood (interrogative, subordinate interrogative, subject *wh*-question, imperative, or declarative), negation (affirmative or negative), and category of main verb (verb with do, verb without do, or modal). We excluded clauses containing modal verbs, subordinate questions, and subject *wh*-questions, which do not permit *do*-support. As with past-tense verbs, we used quantile binning with Laplace smoothing to assign tokens (representing instances where *do*-support was used, or could have been used but was not) to date ranges, setting the number of bins equal to the logarithm of the number of tokens, rounded up. The initial bins across contexts indicate points in time before tokens of *do*-support are observed. The first tokens of *do*-support by context are observed later than Ellegård (negative declarative, 1440; negative question, 1526; affirmative question, 1526 in our data, versus negative declarative, 1425-1475; negative question, 1400-1425; affirmative question, 1425-1475 according to Ellegård). This difference is likely due to biased sampling of innovative or "interesting" authors by Ellegård. We inferred the most likely population size and selection coefficient using `tsinfer`, initializing the Nelder-Mead simplex optimization routine at parameter values $s = 0, N = 500$ (flag `-start:0.0,500.0`).

The difference in FIT *p*-value between negative declaratives (n=14,227) and negative interrogatives (n=606) cannot be attributed to sample size. Among 100 random subsets of 606

tokens of *do*-support in negative declaratives drawn without replacement from the full set of 14,227 tokens, reanalysis yielded a median FIT *p*-value of 0.02.

2.6. Analysis of verbal negation.

We labeled all negated clauses in the Penn Parsed Corpus of Middle English (1100-1500), which contains 52 documents, excluding four documents that contained metered, stilted, or anachronistic writing (cmorm, cmboeth, cmntest, and cmotest): one book of metrical poetry (the *Ormulum*), Chaucer's translation of Boethius' *Consolations of Philosophy*, the New Testament, and the Old Testament. We used CorpusSearch 2 to label sentences for context of negation, clause type, and the presence and positions (preverbal or postverbal) of *ne*, *not*, and *never*. We excluded non-finite clauses, imperatives, questions, constituent negation, contractions, coordinated clauses and negative concord. We classified the resulting 5,475 instances of negative declaratives as pre-verbal, embracing bipartite, and post-verbal sentential negation. We assigned time points by quantile binning with Laplace smoothing as with past-tense verbs. To avoid affects of pseudoreplication within documents, we used six bins so that each bin contains as many distinct documents as possible while reserving adequate power for the FIT. We conducted two tests for selection. For the first transition, we combined the frequencies of embracing and post-verbal negation to test for selection for embracing or post-verbal negation in a background of pre-verbal negation. For the second transition, we combined the frequencies of pre-verbal and embracing negation to test for selection for post-verbal negation in a mixed background of pre-verbal and embracing forms.

CHAPTER 3 : An Exchangeable Model of Social Evolution

The values of many socially-visible traits are arbitrary, such as words (Trudgill, 2004), names (Hahn and Bentley, 2003), decorative styles (Neiman, 1995) or conventions (Wagner et al., 2016), and are thought to evolve neutrally (Hahn and Bentley, 2003; Bentley et al., 2004; Neiman, 1995). Yet neutral evolution (Kimura, 1983) cannot explain (Kandler and Shennan, 2013), for example, why some social traits come and go in cycles of popularity (Ghirlanda et al., 2013) while others become entrenched and markedly persistent (Wagner et al., 2016). We explain these differences in terms frequency-dependent selection in a model of exchangeable competing traits. We infer frequency-dependent fitness relationships from timeseries of first names at birth, registrations with the American Kennel Club, and language spoken most often at home. We find a consistent non-monotonic pattern of frequency-dependent copying rate in first names in more than a century or births and across the United States, France, Netherlands, UK, Canada, and Norway: Names awarded at a rate between roughly 1 in 10,000 and 1 in 1,000 births tend to increase in usage frequency, whereas more rare or more common names decline. Language use in Canada, Mexico and Zambia show the opposite frequency-dependence: Speakers are more likely to adopt a language if it is either common or extremely rare. Dog breed popularity shows strict negative frequency-dependence, with rare breeds always becoming more popular consistent with novelty bias in cycles of fashion. Language use in the United States also shows negative frequency-dependence driven by rarity bias in immigration policy.

Clothing styles come and go, while writing systems outlast empires. Cultural phenemena often involve many arbitrary choices drawn from practically equivalent alternatives. The environment of competition between these alternatives may favor more common or rare variants, even absent any intrinsic differences between them. The influence of commonness or rarity on growth rate in evolution is called frequency-dependent fitness, and it occurs in biological and social contexts. The form this influence takes characterizes the competitive environment and the outcomes of competition. For example, network externalities
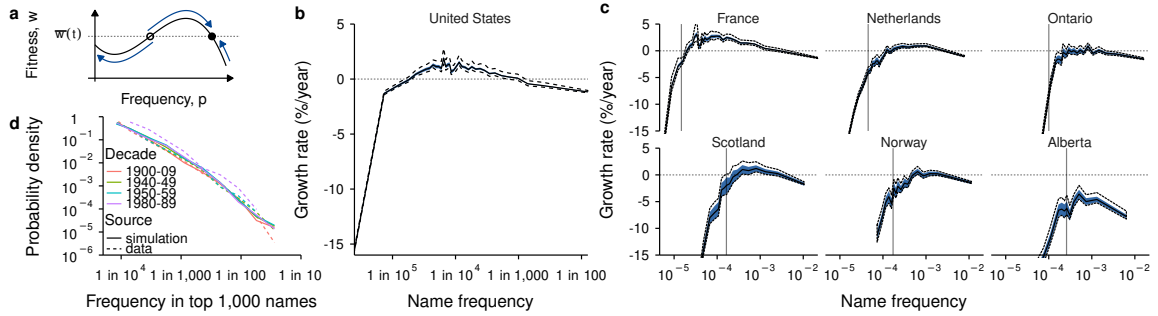
Figure 15: **Frequency-dependent copying of first names.** Fitness is frequency-dependent (a) when the rate of copying a trait depends on the trait frequency. Traits with fitness above the mean fitness, $\bar{w}$, tend to increase (blue arrow), while those below tend to decrease (red arrows), though drift perturbs frequencies in any direction. Equilibria, stable (•) or unstable (◦), exist whenever the frequency-dependent fitness curve crosses the mean fitness. Copying rate of first names depends on the frequency of names in the preceding year in the United States (b), and six other states (c). Panels b-c indicate our inference (solid black, see Methods), bootstrap (conservative, Methods) 95% confidence intervals (dotted lines), and model 95% confidence intervals (blue bands, Methods). Simulations of the model under the inferred parameters reproduce observed frequency distributions of names (d), which are not distinguishable from a neutral copying process (Hahn and Bentley, 2003) due to the strength of drift.

created a monopolizing pressure against the minority videocasette format in the competition between VHS and Betamax (Katz and Shapiro, 1986), whereas balancing selection on a single-locus gene in left- and right-mouthed cichlids (Hori, 1993) favors a 1:1 ratio of the two forms. Frequency-dependence influences diversity, either suppressing rare variants or elevating them to some equilibrium frequency, depending on whether higher abundance leads to faster or slower per-capita growth (Volkov et al., 2005).

Here, we directly infer how the average growth rate (fitness) of a type depends on its frequency (Fig. 15a) from timeseries of competition between types. We model a timeseries as the outcome of a Wright-Fisher process where the fitness of competing traits w is determined only by their frequency x and a trait-independent form of frequency-dependence, $w(x)$, so that the expected frequency of each type i in the next timestep is its current frequency x_i^t

reweighted according to fitness $w(x_i^t)$:

$$E(x_i^{t+1}) = \frac{x_i w(x_i)}{\sum_{j \in \text{types}} x_j w(x_j)}. \tag{3.1}$$

We infer a piecewise-constant approximation to $w(x)$ by quantizing frequency (binning) and computing the maximum-likelihood fitness w_k of types in each frequency bin k (see section 3.1). The inference bears an analogy to image reconstruction (Lange et al., 2000), in which the w_k represent pixel values of an image of $w(x)$. As a consequence of the tradeoff between the number of bins and the number of observations in each bin, more observations allow more bins and progressively better approximations to the true $w(x)$. Thus the piecewise approximation is a non-parametric estimate of the true continuous function $w(x)$.

First names are a classic model system to study the diffusion of cultural traits by random copying (Bentley et al., 2004), as babies are typically given names that are copies of existing names. The US Social Security Administration provides more than a century of high-quality full-count data with minimal censorship. Name frequency distributions are consistent with a neutral model (Hahn and Bentley, 2003) in which selection cannot favor any name, despite that trends in naming have demonstrated psycholocial biases (Berger et al., 2012; Coulmont et al., 2015). We expect that when choosing a name for a newborn child, a parent may wish to avoid rare or invented names, which could provoke ridicule, but also avoid very common forms, which might seem mundane. Biases of this sort would induce frequency-dependent selection in name copying.

We infer the curve of frequency-dependence in an annual timeseries of first names of United States Social Security Card recipients born in the United States since 1880, which censors only names represented in less than 5 births in a given year (Fig. 15a). Most names are extremely rare. While only 275 names ever reach the highest frequency bin (greater than 1 in 300), more than 100,000 names are found in less than 1 in 190,000 births in some year. Among these rare names there are on average 0.84 copies per capita the next year. Yet the commonest names do not fare well either, leaving on average 0.9883±0.0000 copies (95%

CI). Meanwhile, the highest fitness names, which leave 1.019±0.002 copies per year, are any names around 1 in 15,000 births. The inferred $w(p)$ indicates that the per-capita growth rate of a name depends on its frequency among last year's births, and therefore parents in the United States tend to award names that are neither too common nor too rare at the time.

Furthermore, a similar pattern is visible in five other states that publish adequate records of first names (Fig. 15b): France, the Netherlands, Ontario, Scotland, Norway, and Alberta, in order of decreasing population size. Scotland, the Netherlands, and Alberta publish a complete list of names with no censorship of rare names, allowing us to observe the fitness of names of which there may be as little as one copy per year. In each case, the highest average growth rates occur for names awarded between 1 in 990 and 1 in 30,000 names. The highest growth rate is positive in all states except Alberta, due to the high rate of novel names in Alberta. Alberta also experienced drammatic population growth and immigration: the population has grown 19%/decade on average since 1981, of which 72% is due to immigration (34% domestic and 38% international). High immigration implies that a higher growth rate is required to achieve a replacement rate of copying in the face of novel incoming names, and therefore even the highest growth rates of names in Alberta are not enough to replace themselves.

Previous studies of rank-abundance distributions (Hahn and Bentley, 2003) did not reveal selection among names. We replicated these results by simulating the model with the inferred parameters (Methods) and found no significant difference between rank-abundance distributions derived from the true data, simulations under the inferred parameters (Fig. 15d), or simulations of the neutral model. This indicates that frequency-dependent growth, when it is present, might not be evident in the rank-abundance distribution, and so the form of frequency-dependence is not in general identifiable from rank-abundance data. We are able to reject neutrality in name copying only by using much more information present in timeseries.
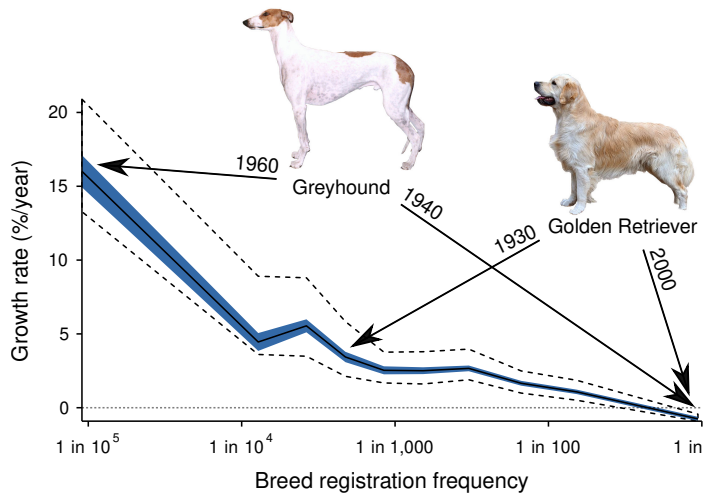
Figure 16: **Frequency-dependence in American Kennel Club registrations.** Popular dog breeds are subject to fads. We infer the average frequency-dependence in the rate of registration with the American Kennel Club among 153 dog breeds from 1926 to 2005. While the rarest recognized breeds increase in registration frequency at 16%/year, a registration among the commonest breeds inspires only 0.992 registrations of that breed in the next year, on average.

Dog breed registrations have been studied to understand cycles of fasion (Ghirlanda et al., 2013). In fashion and art markets, more common goods are often valued less, or new, rare goods valued more O'Dwyer and Kandler (2017). Indeed, we infer negative frequency-dependent growth rates of registrations by breed in an annual timeseries (1926-2005) of registrations of the American Kennel Club (Fig. 16). In both first names and dog breeds, the most popular types are selected against, and timeseries of popularity show clear, smooth boom-bust cycles. Boom-bust cycles or a fitness benefit to novel types are impossible under the exchangeable model. However, the form of frequency-dependence inferred under the exchangeable model also estimates the average fitness as a function of frequency at equilibrium. Novelty bias creates effective negative frequency-dependence, and so the negative-frequency-dependent $w(p)$ we observe for dog breeds and common names is likely due to preferences for novel types. Furthermore, whatever the detailed mechanisms, negative frequency-dependent effects broadly encourage diversity while positive frequency-dependent effects suppresses it.

Positive frequency-dependence causes standardization (Wagner et al., 2016) or regulariza-

tion (Pagel et al., 2007; Ferdinand et al., 2013). Language is the prototypical social standard: most languages express most concepts just as well as any other, yet the utility of speaking a particular language depends strongly on social context. Are people more likely to adopt more common languages? If so, an increasingly interconnected global society would pose a direct threat to all but the most common languages, as has been widely reported (Krauss, 1992; Mufwene, 2002; Abrams and Strogatz, 2003).

We directly examined the influence of commonness or rarity on language shift using the framework of frequency-dependent selection. Language shift occurs for many reasons that might relate to culture, economic access, politics, history, or identity (Mufwene, 2001). Nonetheless, imposing a model of frequency-dependent selection isolates the average effect of frequency on language-shift, which has often been of interest (Crowley, 1995; Abrams and Strogatz, 2003).

Canada's rich history of multilingualism and self-reported census data on language spoken at home provide a testbed for answering the question. We inferred the frequency-dependent fitness relationship using a four-point timeseries of 112 language categories for roughly 30 million speakers in each 5-year Census of Population since 2001 (2001-2016, Methods, Fig. 17). Our measurement of frequency-dependent growth rate conflates growth in language use due to language shift, immigration, and human demography into a single measure of per-capita propensity to reproduce one's language as a function of the language's frequency among speakers. Most languages (63%) changed frequency bin during the timeseries, and all non-empty bins included at least six languages except for the highest-frequency bin, which contained only English and French. Consistent with naïve expectation, average growth rate increases with commonness from around -5.9%/year for languages represented by around 1 in 100,000 speakers to 0.1%/year for Canada's most frequent languages, English and French. Yet the highest growth rates are found among the rarest languages, spoken by around one in one million—languages spoken by 10s to 100s of people. Among these rarest languages, we find Celtic, which is globally endangered (Kandler et al., 2010), and languages
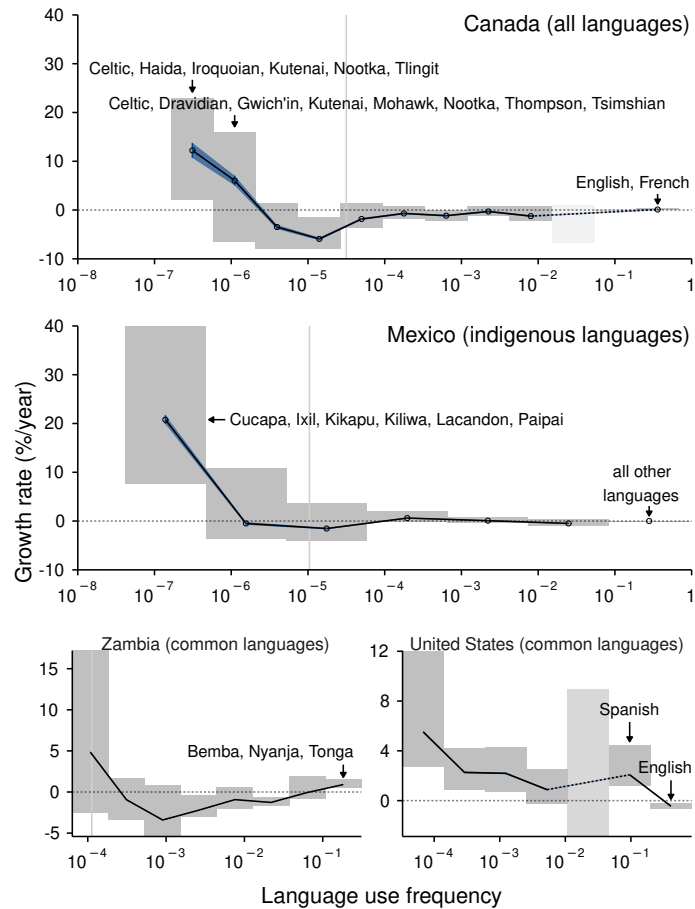
Figure 17: **Frequency-dependent bias in language shift.** The rarest languages are more likely to be adopted per capita than more frequent languages among all languages in Canada (112 languages), indigenous languages of Mexico (57), and common languages in Zambia (61) and the United States of America (US; 84). Selection coefficients are inferred for logarithmic frequency ranges from national census data on language spoken at home (Canada: 5-year census, 2001-2011; Mexico: 10-year census, 1990-2010; Zambia: 10-year census, 1990-2010; US: American Communities Survey 1-year estimates, 2000-2015). Labels with arrows list all languages whose frequency was ever within the indicated interval. Blue bands in each plot indicate model-predicted variability in parameter estimates (95% confidence intervals). Vertical gray lines indicate the frequency corresponding to 1000 speakers. Grey boxes indicate sensitivity to resampling underlying data (middle 95 quantiles of bootstrap), shaded according to the number of bootstrap runs which included languages in that frequency range. The lowest frequency interval inferred from US data is not shown due to problems with partial censorship in the source data.

of the Dravidian family, rare in Canada but common in parts of India and Sri Lanka. The preponderance, however, are indigenous (Gwich'in, Haida*, Iroquoian, Kutenai*, Mohawk, Nootka/Nuučaanuł**, Thompson/Nlaka'pamuctsin*, Tlingit**, and Tsimshian**), of which most are native to the Pacific coast. These cultures underwent brutal repression persisting to the end of the 20th century that curtailed use of their languages by disrupting the transmission of culture across generations (Truth and Reconciliation Commission of Canada, 2015). The majority of these languages are now considered severely (*) or critically (**) endangered (Moseley, 2010). Previous models based on demography have reported and predicted precipitous declines (Krauss, 2007). The contrary high growth rate in adoption we observe could be attributable to confounding factors, to historical rebound, or to intrinsic benefits to speakers of rare languages.

We replicated our analysis in three timeseries from Mexico, Zambia, and the United States, to understand the generality of our finding and control for confounding factors. In Mexico, the decennial census specifically measures the use of indigenous languages, as the government must provide services in indigenous languages by Article 2 of its constitution and the Ley General de Derechos Lingüísticos de los Pueblos Indígenas (General Law of Linguistic Rights of Indigenous Peoples) of 2003. The decennial census provides a three-step timeseries (1990-2010) of 5-7 million speakers in roughly 100 language categories, some with as few as 4 speakers. The growth rates we measure from this timeseries conflate only demograpy and language-shift, thus providing a natural control for effects of immigration, as indigenous speakers can be assumed not to immigrate. Indeed, indigenous languages of Mexico replicate the pattern seen in Canada of the highest growth rates in the rarest languages, so we conclude that this pattern is not due to immigration.

In Zambia, detailed decenial census data is available on the use of 61 languages over approximately 10 million speakers and 30 years (1990-2010). Zambia's linguistic diversity likely far exceeds what was recorded, as the languages of 50,000 speakers were not represented in the 61 recorded languages, and the lowest recorded frequency was 320 speakers of Mbowe.

Despite being a British colony before its independence in 1964, English, Zambia's official language, is spoken by a small minority (1.1-1.5%), reflecting how language adoption depends on social and economic details despite broadly similar patterns of colonization (Mufwene, 2001). Our inference of frequency-dependence in Zambia (Methods) is consistent with both the highest growth rates in the rarest languages, and the pattern of increasing growth rate from intermediate to the most common languages. Furthermore, self report census data has been criticized due to the incentives of respondents to exagerate or conceal rare language use for political or historical reasons (Krauss, 2007). Yet, our inference of the same pattern across three countries and two continents with very different histories, cultures, and social and economic standing of minorities shows that biases in self report data is not influencing our result.

In the United States, the American Communities Survey provides an annual timeseries (2000-2015) of the useage rates of 89 common language categories as public-use microdata. Our analysis (Methods) reveals a general pattern of negative frequency-dependent growth rate across the range of frequencies present in the data. In particular, English, the only representative of the highest frequency category, is below replacement fitness, declining at a rate of -0.4%/year, consistent with US Census reports on increase in the use frequency of languages other than English over recent decades driven in large part by immigration (Ortman and Shin, 2011). More generally, the growth rates of non-English speach communities in the United States are controlled by immigration. Immigration in the United States is strictly regulated with caps on numbers of immigrants from particular countries. These caps limit the most common immigration flows, and thus inherently impose negative frequency-dependence. We therefore conclude that our observation of negative frequency-dependence is primarily to immigration, and thus cannot measure language shift.

A movement within linguistics now more than 25 years old (Krauss, 1992) has sought to advertise (Nettle, 2000; Miyaoka et al., 2007) and even reverse (Fishman, 1991) the decline of rare languages worldwide. Our findings run contrary to the percieved instability

1938 Name	Current Name	1938 Ct.	Cur. Ct.	Current Source
Canada				
Haida	Haida	2,500	10	2011 Census
Iroquois	Mohawk	8,000	155	2011 Census
Tukudh	Gwich'in	2,000	60	2011 Census
United States				
Arapahoe	Arapaho	2,000	1,065	2009-2013 ACS
Arikara	Arikara	400	155	2009-2013 ACS
Cheyenne	Cheyenne	3,000	1,920	2009-2013 ACS
Hidatsa	Hidatsa	470	625	2009-2013 ACS
Hopi	Hopi	4,000	6,075	2009-2013 ACS
Keres	Keres	5,000	13,190	2009-2013 ACS
Seneca	Seneca	3,000	1,015	2009-2013 ACS
Shawnee	Shawnee	1,000	255	2009-2013 ACS
Winnebago	Winnebago	1,000	1,720	2009-2013 ACS
Mexico				
Zapotec	Zapoteco	5,000	477,070	Census 2010
Northern Rhodesia				
—	—	—	—	—

Table 4: **Languages with less than 10,000 speakers prior to 1938 in Canada, United States, Mexico, and Northern Rhodesia.** 1938 values are all languages that appear in *The Book of a Thousand Tongues* (North, 1938) in the geographic index for the respective country with a number of speakers less than 10,000 indicated. All 15 languages listed for Northern Rhodesia had no number of speakers indicated or a number greater than 10,000. Current counts are based on questions about language spoken at home in Canadan and US census data, or indigenous language spoken in Mexican census. Although the 1938 name "Iroquois" could refer to any one of several languages, the translation listed is Mohawk, and its translator, Joseph Onasakenrat (1845-1881), was a Mohawk-speaking Chief who lived and attended seminary near Montreal, and so this is presumably the population of "some 8,000 Indians, living in Canada, just north of Montreal" (North, 1938).

of rare languages. In comparison to the predictions of neutral models used in language contact (Trudgill, 2004) and elsewhere (Lycett, 2008), the negative frequency-dependent growth rate we find predicts longer persistence and lower probability of extinction for rare languages. Furthermore, models that use age and geography of native speakers (Krauss, 2007) to avoid questions about the legitimacy of self report data have done so at the cost of constitutively denying the possibility of language uptake by younger generations. We therefore question whether rare languages are fragile at a baseline. Linguists have reported that languages with less than 10,000 speakers are in danger of extinction the timescale of a century (Dixon, 1991; Crowley, 1995). *The Book of a Thousand Tongues* (North, 1938), a

forebearer to the *Ethnologue* (Canonge and Pittman, 1958) index of world languages, lists 13 languages with fewer than 10,000 speakers prior to 1938 in Canada, the United States, Mexico, and Northern Rhodesia. Of these, all still appear in current national censuses (Table 4). Following smallpox epidemics of 1860, the population of Haida speakers was reduced to around 450 (Van Den Brink, 1974), yet it is still spoken and taught to children today (Cf. e.g. the Skidegate Haida Immersion Program, HlGaagilda Xaayda Kil Naay). Furthermore, inappropriately treating rare languages or cultures as fragile of themselves ignores potential acute threats such as genocide, extractive economic projects, occupation, or religious or national indoctrination programs.

The high growth rate of rare languages is encouraging for rare language preservation, but should not be taken as relief from concern. Languages with fewer speakers are still presumably nearer to extinction. Furthermore our model predicts these same languages to be subject to much stronger stochastic drift than common languages, and so still suffer a high danger of local extirpation or extinction. Language shift is a complex topic with many determining factors unique to particular languages such speakers' exposure to the global economy (Mufwene, 2002), yet some component of language growth or decline is due to the size of its pool of speakers.

Inferring frequency-dependent selection reveals something about the competitive environment of alternative equivalent traits. Although we have examined social traits here, genetic and phenotypic timeseries in biology are becoming more available and likely to reveal the ecological processes underlying evolutionary change.

Our model is exchangeable in the sense that no particular trait value is treated differently: all experience the same frequency-dependent selection. Nonetheless, frequency-dependent growth explains why one element, such as VHS, may remain highly popular even when alternatives are intrinsically superior. The exchangeable model generalizes neutral models by allowing selection without special treatment for any particular type. On the other hand, the exchangeable model is a simple special case of frequency-dependent selection where each

type is allowed to have unique behavior. The general case of frequency-dependent selection is notoriously complex (Nahum et al., 2011; Nowak and Sigmund, 2004), but the exchangeable model provides a tractable special case. Classic cases of frequency-dependent selection such as sex ratio evolution (Karlin and Lessard, 1986) and handedness in scale-eating cichlids (Hori, 1993) fit neatly within the exchangeable model.

3.1. Inferrence

We infer an effective frequency-dependent selection in a Wright-Fisher process. We assume that the relative fitness of a type at frequency p is $w(p)$. The Wright-Fisher process allows us to write the transition probabilities in the infinite alleles process with exchangable alleles of fitness $w(p)$ explicitly, and thus simulate the process.

If we can describe the true $w(p)$ using a few parameters, the method of maximum likelihood allows us to infer the parameters if we can optimize the likelihood of a timeseries. The transition probabilities in the Wright-Fisher process are the multinomial sampling probabilities from the current frequencies, weighted by the marginal fitness:

$$X_{t+1}|X_t \sim \text{Multinom}(N_t, w(X_t) \cdot X_t). \tag{3.2}$$

Thus, the probability of any sample path $X_0, ..., X_t$ is the string of conditional probabilities $\Pr(X_t|X_{t-1}) \cdot \Pr(X_{t-1}|X_{t-2}) \cdot ... \cdot \Pr(X_1|X_0)$. Each conditional probability follows the multinomial distribution function, $f(x, \pi)$. Viewed as a function of the observation x, f is the distribution function, whereas as a function of the parameters π, f is the likelihood function. Thus, given a realization of the standards process x_t, $\prod_{i=1}^{t} f(x_t, w(x_{t-1,i}, \pi)_i \cdot x_{t-1})$ gives the likelihood of the parameters π. Knowing the form of $w(x, \pi)$, such as $w(x, \alpha) = 1 + \alpha x$, we can compute the likelihood of a given α. We use the GNU Scientific Library provides to compute the log-probability (log-likelihood) of each multinomial sample given the expectations derived from $w(p)$.

Finding the optimal parameter set, in particular when $w(p, \theta)$ may have many parameters,

may be challenging. When the log-likelihood is concave, better optimization algorithms are available. The likelihood of a sample path is the product of the conditional likelihoods, and therefore the sample path log-likelihood has the same concavity as each conditional likelihood. The multinomal log-likelihood is

$$\mathcal{L} = \ln N! - \sum_i \ln x_i'! + \sum_i x_i' \ln \pi_i(x, \theta), \tag{3.3}$$

where π_i refers to probability a random birth is of type i. Writing this in terms of w, and bundling terms that are not a function of the parameters into a constant c,

$$\mathcal{L} = c + \sum_i x_i' \ln \frac{x_i w(x_i, \theta)}{\sum_j x_j w(x_j, \theta)}$$

$$= c + \sum_i x_i' \ln x_i w(x_i, \theta) - \sum_i x_i' \ln \left(\sum_j x_j w(x_j, \theta) \right)$$

$$= c - N \ln \left(\sum_j x_j w(x_j, \theta) \right) + \sum_i x_i' \ln x_i w(x_i, \theta).$$

If $w(x, \theta) = 1 + \theta x$, for example this would be

$$\mathcal{L} = c + \sum_i x_i' \ln \frac{x_i(1 + \theta x_i)}{\sum_j x_j(1 + \theta x_j)}. \tag{3.4}$$

To find local maxima of the likelihood, we would like to take the derivative with respect to θ.

$$\frac{d}{d\theta} \left(-N \ln \left(\sum_j x_j w(x_j, \theta) \right) + \sum_i x_i' \ln x_i w(x_i, \theta) \right)$$

$$= -N \frac{\sum_j x_j \frac{d}{d\theta} w(x_j, \theta)}{\sum_j x_j' w(x_j, \theta)} + \sum_i x_i' \frac{\frac{d}{d\theta} w(x_i, \theta)}{w(x_i, \theta)}$$
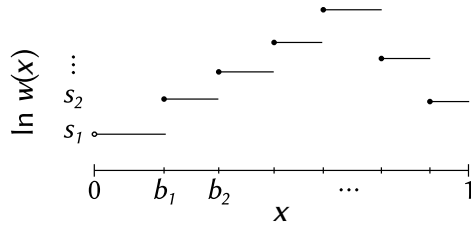
Figure 18: We let $w(x)$ be a piecewise constant function with discontinuities at b_i, and selection coefficient s_i within each bin.

with extrema found at

$$\sum_i x_i' \frac{\frac{d}{d\theta} w(x_i, \theta)}{w(x_i, \theta)} = N \frac{\sum_j x_j \frac{d}{d\theta} w(x_j, \theta)}{\sum_j x_j w(x_j, \theta)}.$$

We produce a non-parametric approximation to the function $w(x)$ by dividing frequencies x into bins. That is, consider the case when w is a piecewise-constant function of x, with parameters $\theta = \{s_i\}$ for each piece. We then infer the most likely fitness of types within each frequency interval under the model using methods from image reconstruction. This gives a discrete frequency-dependent fitness function, which is a good approximation to the true continuous function. Using finer intervals gives a better approximation until too little data is available to get precise parameter estimates for each interval, analogous to the tradeoff between spatial precision and noise in images.

The full model, is then a Wright-Fisher process on a population $N(t)$ with mutation strength $N\mu$ and a piecewise-constant realtive fitness function $w(x|s_k)$ with D bins with boundaries $b_1, ..., b_{D-1} \in (0, 1)$ and fitnesses $s_k, k \in [1, ..., D]$ where $w(x) = e^{s_k}$ when $x \in (b_{k-1}, b_k]$. For notational simplicity we let $b_D = 1$ and $b_0 = 0$, so that $(b_0, b_D] = (0, 1]$. Frequency 0 is not in the domain of $w(x)$: types that were at frequency 0 in the last generation are considered mutants introduced at mutation rate μ. The likelihood of an update to counts x_i' of each

type i and m new mutants from current counts x_i is given by

$$\mathcal{L}(x_i|s_k, \mu, N) = \ln N! - \ln m! - \sum_i \ln x_i'!$$

$$+ m \ln \mu + (N - m) \ln(1 - \mu)$$

$$+ \sum_i x_i \ln \left[x_i' \frac{x_i}{N} w(x_i) \right] - (N - m) \ln \left[\sum_i \frac{x_i}{N} w(x_i) \right] \qquad (3.5)$$

Maxima of the likelihood are obtained by setting the derivatives with respect to all parameters equal to zero. This results in a system of equations. The equation for μ does not involve any other parameters and thus μ can be estimated separately.

$$\frac{\partial}{\partial \mu} \mathcal{L} = \frac{m}{\mu} - \frac{N - m}{1 - \mu} = 0, \qquad (3.6)$$

which implies $m = N\hat{\mu}$: the best estimate of the mutation rate is the observed fraction of mutants. Over the timeseries then, $\frac{\partial}{\partial \mu} \sum_{t=1}^{T} \mathcal{L}_t$ implies $\hat{\mu} = \sum_{t=1}^{T} m_t / \sum_{t=1}^{T} N_t$: the best estimate of per-generation mutation rate is the observed fraction of mutants over the entire timeseries. The condition for maximum-likelihood $\hat{s}_k$ is the set of equations $\frac{\partial}{\partial s_k} \mathcal{L} = 0$ for all $k \in [1, ..., D]$. This amounts to the system of simulaneous equations, one for each k,

$$\sum_{\substack{t \in [1, T-1] \\ i:x_i^{(t)} \in (b_{k-1}, b_k]}} x_i^{(t+1)} = \sum_{t \in [1, T-1]} \left[(N_{t+1} - m_{t+1}) \left(\frac{\sum_{j:x_j^{(t)} \in (b_{k-1}, b_k]} e^{\hat{s}_k} x_j^{(t)}}{\sum_j e^{\hat{s}_k(x_j/N_t)} x_j^{(t)}} \right) \right]. \qquad (3.7)$$

This statement has the interpretation of "observed equals expected": the left hand side is the observed counts emmanating from the kth bin and the right side is the expected count given the population configuration at time t and number of mutations that occurred between t and $t + 1$. That is, setting the expected data equal to the observed data maximizes the likelihood.

The system of equations (3.7) is difficult to solve, numerically or otherwise, primarily because of the $\hat{s}_k$ that appear in the exponent in the sum in the denominator on the right

hand side, which typically makes the equation for each $\hat{s}_k$ non-linearly dependent on all others.

To solve the equation, we use an MM or "Minorize and Maximize" algorithmic strategy (Lange et al., 2000). The overall strategy is to replace the likelihood $\mathcal{L}(s)$ with a family of surrogate minorant functions $g(s|s_m)$ that have two properties: $g(s_m|s_m) = \mathcal{L}(s_m)$, and $g(s|s_m) \leq \mathcal{L}(s)$, for whatever s, s_m are in the domain of $\mathcal{L}$. The minorant is guaranteed to be at most $\mathcal{L}$ everywhere and to match $\mathcal{L}$ and its derivative at some point s_m. One can reason that s_m is either the optimum of $\mathcal{L}$, in which case the derivative of g and $\mathcal{L}$ at s_m is zero, or else, providing g and $\mathcal{L}$ are smooth functions, the derivative of g is non-zero and the optimum of g likes somewhere between s_m and the optimum of $\mathcal{L}$. If we chose a g that is easy to optimize, we can monotonically approach the optimum of $\mathcal{L}$ by finding the optimum s in $g(s|s_m)$, and then using that s as s_m in the next iteration. Each iteration closes the gap between s_m and the true optimum of $\mathcal{L}$.

To accomplish this, we must choose an appropriate minorant. Many minorants are possible, such as a quadratic approximation, as discussed in Lange (2010), and I investigated a few before finding the right one. Examining the likeihood function Eq. 3.5, we can notice that the very last term,

$$-(N-m)\ln\left[\sum_i \frac{x_i}{N} w(x_i)\right]$$

involves a sum inside the log, which gives rise to the troublesome denominator. This function in turn, since it is a negative log, is easily minorized using the term's Taylor expansion:

$$\ln\left[\sum_i \frac{x_i}{N} w\left(x_i|s^{(m)}\right)\right] + \sum_{k=1}^{D}\left(s_k - s_k^{(m)}\right)\frac{\partial}{\partial s_k}\ln\left[\sum_i \frac{x_i}{N} w\left(x_i|s^{(m)}\right)\right] \geq \ln\left[\sum_i \frac{x_i}{N} w(x_i|s)\right]$$

The derivative with respect to s_k (though not $s_k^{(m)}$) on the left-hand is linear, whereas the derivative with respect to s_k of the right-hand side, as we have seen, is not. Setting the derivatives of the minorant to zero yields a system of equations identical to Eq. 3.7 except that the $\hat{s}_k$ in the denominator has been replaced by $s_k^{(m)}$. Now the kth equation

is a function only of s_k and the optimization is trivial. I use these optima as the new s_m and proceed until no progress is achieved between iterations, that is, the derivative of the minorant at s_m has reached approximately zero. At this point, the minorant and $\mathcal{L}$ are both maximized.

Two facts are notable about the maximum-likelihood parameters. First, the mutation rate is independent of the parameters s_k. This is important since the mechanisms of mutation are often unknown or ill-defined in cutural contexts. For example, in order for a name not present in the previous year to appear in the US Social Security Administration database, it must appear in 5 births in that year. Thus our measurement of mutation rate conflates the probabilities of an initial mutation and a subsequent rise to count 5. This results in a biased (unfairly low) estimate of mutation rate, but is irrelevant to the parameters s_k as they are independent. Second, the derivatives of likelihood do not depend on N except where N normalizes counts to frequencies. Thus, the inferrence depends on the frequencies of types, but not the total population size: The variance in frequency increments are ignored. This indicates that estimates of s_k and estimates of effective population size (by whatever means) are also independent. Thus, changing timescales or population sizes (while suitably transforming units) has no effect on the inferrence of s_k in the diffusion limit assuming time intervals are short. Alternatively stated: if each increment of the data were replaced by the same increment followed by n generations of neutral Wright-Fisher evolution, the expected value of the estimators of s_k are unchanged.

3.2. Replacement fitness calculation

In Figure 17 and elsewhere, we compute the replacement fitness, $\bar{w}$. The replacement fitness is a kind of average fitness. It is the fitness, which, if a type with that (constant, frequency-independent) fitness were born at the beginning of the timeseries, its expected frequency at the end of the timeseries would be the same. The time-weighted population average fitness for generations $t \in \{1, ..., T\}$ and frequencies $x_{i,t}$ of types $i \in \{1, ..., k_t\}$ at generation t, all

84

types having frequency-dependent fitness $w(x) = e^{s(x)}$, is

$$\bar{s} = \left(\frac{1}{T}\right) \sum_{t=1}^{T} \log\left(\mu + \sum_{i=1}^{k_t} x_{i,t} w(x_{i,t})\right).$$

3.3. Confidence intervals

We use two confidence intervals (CIs) in this study: a model-based CI and a data-based CI. The model-based CI is an accurate theoretical prediction of the variance on the estimator subject to the assumption that the data is generated by the model. The data-based CI is a non-parametric bootstrap (Efron and Tibshirani, 1986) which generates an empirical distribution of the estimator by resampling the underlying data. The data-based CI makes no assumption that the data are generated from the true model, but does assume that a resampling scheme is available that generates equally probable samples, such as multinomially resampling the observed dataset. Both of these assumptions may fail. Corroboration between the two CIs, as we observe in baby names and dog breeds, supports the accuracy of the model and the underlying assumptions. When the different CIs disagree, the data-based CI is more conservative, as it accounts for failures of the model to describe the data.

For the large samples used to infer dog breeds and baby names, we use model-based confidence intervals on parameter inferences computed using the observed information (Efron and Hinkley, 1978), a measure of the curvature of the likelihood function at its maximum. In the limit of may independent observations and an identifiable inference, the log-likelihood surface is approximately parabolic near its maximum. The observed information is the matrix of second partial derivatives (Hessian matrix), and in this limit, its inverse is the variance-covariance matrix of the estimators $\hat{s}_k$. We compute 95% confidence intervals on each paramter as $1.96\sigma_{\hat{s}_k}$ by projecting the variance-covariance matrix onto each axis k. We verified the confidence intervals by simulation using a parametric bootstrap. By simulating equal sample sizes to our data using the model with the inferred parameters, we find that the 95% confidence intervals derived from the observed information include the

true parameters within statistical error of 95% of the time after 1000 trials for each of our datasets.

We compute data-based CIs by resampling transitions within the timeseries. Bootstraps that accomodate the dependence relationships that may be present in timeseries is an open-ended topic (Politis, 2003), because generally there is no resampling scheme available that logically guarantees that the resampled data is equiprobable under the true generating process, in contrast to the boostrap argument for independent samples. Nonetheless, in our case, the likelihood calculation sums over observations minus expectations of change from one timestep to the next, and so it is these transitions that are assumed to be independent between types and across time periods. Unaccountedfor dependence between observations enters by either failures of the Markovian assumption in the true generating process, or by effects of competition between types induced by finite population size. We attempt to resample these in a way that preserves independence, but also roughly preserves population size, in order to preserve the meanings of frequency. We resample a given timestep by chosing at each timestep transitions $x_i \rightarrow x_i'$ from that timestep in the data uniformly with replacement. We determine the number of transitions to include at each timestep by iteratively including another transition sample with probability $\max(1, p - c/e - c)$ where c is population size of thus-far-included transitions, p is the observed population size in the original data, and e is the expected population size increase by including one more transition sample. This guarantees that the expected resampled population size is equal to the observed population size. This resampling scheme is equivalent to a multinomial resampling of independent observations which accounts for robustness of the estimator to sampling noise, but also preserves the frequency interpretation of the original count data. The frequency of a given count is allowed to vary around its expectation in the bootstrap, and thus the bootstrap also represents robustness of the estimator to the inclusion or omission of observations due to censorship in the data, since such omissions would cause the frequencies in observed data to fluctuate. As in the true data, types of high frequency are more likely to be observed. Our bootstrap thus represents robustness of the estimator to

sampling noise as well as censorship bias against low-frequency types. We use data-based CIs for the small samples available for language use. Model-based CIs are not accurate in these cases due to non-asymptotic behavior, nor does the model fully account for the data. The bootstrap CIs are conservative estimates of the true 95% CI.

3.4. Inferring N_e

Our primary concern thusfar has been estimating fitness as a function of frequency, which determines the mean change (advection) in the diffusion equation. The diffusion term is also of interest. In the diffusion limit of the Wright-Fisher process, the diffusion term is $(1/2)\partial^2/\partial x^2 (x(1-x)/N)\phi(x,t)$, where N is the number of allele copies in the Wright-Fisher process (Kimura, 1964), and $\phi(x,t)$ is the probability density at frequency x at time t. The coefficient $x(1-x)/N$ describes that variance in gene frequency accrues at a rate of one binomial sampling per generation, and has the units of a transport coefficient: frequency^2/time. The parameter N represents the Wright-Fisher effective population size, and is equal to the census population size (or its geometric mean if it fluctuates) when one round of random assortment (a single multinomial sample) occurs at each sampling interval, g, i.e., one generation. For short time intervals, variance accrues linearly in time in the diffusion equation, and so if sampling is conducted at a different interval g', N' will be rescaled proportionately such that $g/g' = N'/N$. Thus, at sampling interval g and census population size N, if the inferred effective population size N_e differs from N, then there is an effective generation time $g_e = gN/N_e$ which is the time for the diffusion to accrue the same variance as one Wright-Fisher generation—one multinomial sample of the full census population.

We infer N_e by rescaling frequency increments in order to produce homoscedastic updates under the inferred model parameters. If r_i are the residuals of the data, $r_i = x_i - e_i$, where x_i are the data and e_i are the expectations under the model, then the rescaled increments

are

$$Y_i = \frac{r_i}{\sqrt{2x_i(1 - x_i)}}. \tag{3.8}$$

The rescaled increments Y_i have variance $1/N_e$ per generation, so $N_e = 1/S$ where S is the sample variance of the Y_i.

We estimate the effective population size N_e and a generation time for baby names in each country in Table 5, where the generation time represents the characteristic timescale on which complete mixing of names occurs.

Country	Years	Births/year	Inferred N_e	$N_{e,s=0}$	Gen. time N/N_e	μ	Censorship
United States	1880–2015	1,818,516	371,920	386,140	4.89	0.007	< 5 births/year
France	1900–2015	676,734	48,882	53,260	13.84	0.006	< 4 births/year or < 20 ever
Netherlands	1946–2014	220,590	90,764	137,814	2.43	0.067	None
Ontario	1917–2013	99,698	23,680	25,172	4.21	0.019	< 5 births/year
Scotland	1974–2015	60,921	23,506	33,770	2.59	0.053	None
Norway	1946–2015	58,196	19,073	19,486	3.05	0.011	< 4 births/year
Alberta	1980–2015	38,607	796	1,187	48.52	0.145	None
Ireland	1998–2015	43,164	3,756	3,707	11.49	0.012	< 3 births/year

Table 5: Census population size N (geometric mean number of births per year) and Wright-Fisher effective population size N_e inferred from the model residuals after fitting over the duration of each timeseries. The innovation rate μ is the per-capita per-generation rate that a novel name reaches observable frequency. The minimum observable freqeuncy varies by country due variable policy regarding censorship of small counts as identity protection, thus μ is lower in countries with more censorship, due to the decreased probability of achieving observable frequency before extinction. Inferred $N_{e,s=0}$ is the N_e inferred from the residuals of the neutral Wright-Fisher process.

3.5. Known model misspecification biases

Inference of parameters by maximum likelihood is unbiased when observations originate from the model assumed in the likelihood calculation. Real processes never conform perfectly to a probabilistic model, so model misspecification can nonetheless introduce bias. One source of model misspecification is that we believe the true process (births or adoption of languages) to be continuous in time, but data is only available in discrete time. Hence, some temporal variation in the frequency trajectories goes unobserved, leading to bias towards neutrality. In the discrete-time process, a type at one frequency $s(p_1)$ moves to another frequency in the next timestep p_2 with probability depending on $s(p_1)$. In the

continuous time process, however, the frequency $p(t)$ may visit many frequencies in the neighborhood of p_1 and experience different fitness $s(p(t))$. Temporal discretization removes the record of $p(t)$, and so we attribute all selection experienced in bringing a type from frequency p_1 to p_2 to its fitness at p_1, $s(p_1)$. This averages s between adjacent segments of the piecewise-constant frequency-dependent fitness, creating a bias that tends to equalize fitnesses between adjacent parameters. As a result, fitness values may be more extreme (further from 0) than what we report.

3.6. Binning

The bin boundaries b_k in the parameterization are arbitrary. The size or number of bins represents a tradeoff between resolution in frequency and resolution in growth rate.

We employ two binning schemes: log-evenly-spaced binning, and quantile binning. In quantile binning, we chose a number of bins to represent a subjective optimum between frequency and growth rate accuracy. We then assign bin boundaries so that equal numbers of transitions in the data fall within each bin, thus the bin boundaries are the quantiles of the frequencies that occur in the data. In log-evenly-spaced binning, we chose a number of bins, and the bin boundaries are chosen to evenly partition (in log space) the range of frequencies present in the data.

3.7. Particulars

3.7.1. Language spoken at home

We use census data that is designed to reflect language spoken at home for four countries.

For Canada we use tables of "Detailed Language Spoken at Home" or "Detailed Language Spoken Most Often at Home" from Census 2001, 2006, 2011, and 2016. Each census reported a different set of language categories. We harmonized this four-point timeseries by rewriting, merging, or excluding language categories that changed over the course of the timeseries. When langauge category names changed cosmetically, we rewrote them to a single category,

e.g. interpreting "Nootka (Nuu-chah-nulth)", "Nuu-chah-nulth (Nootka)" and "Nootka" as "Nootka" for all timepoints. We merged categories when one category present in a (typically earlier) timestep included multiple categories present in other (typically later) timesteps, and the languages involved are similar, e.g., we aggregated "Cree", "Plains Cree", "Swampy Cree", "Woods Cree", "Northern East Cree", "Southern East Cree", "Moose Cree", "Michif", "Oji-Cree", "Cree, not indicated elsewhere", and "Cree, not otherwise specified" into the 2001 language category "Cree" in each timestep. We excluded language categories which were too general to indicate which language was spoken (e.g. "Creoles", "Chinese, not otherwise specified" or "Austro-Asiatic Languages, not indicated elsewhere") or which were recorded in only some census timepoints (e.g. "Burmese"). The full mapping between categories is given in Table 6.

We inferred the frequency-dependent growth rate of languages in Mexico from a timeseries of the indigenous languages (lengua indígena) from decennial national census 1990-2010 accessed through IPUMS International (Ruggles et al., 2003, Harmonized variable LANGMX with Person weight). We merged subcategories of Zapoteco, Mixteco, Pame, Chinanteco, and Pima because Census 2010 did not record such subcategories, and excluded languages which were not recorded in each census. We assume that non-indigenous languages share the same, constant average fitness. As there is certainly variation in frequency and fitness in non-indigenous languages, our assumption is equivalent to assuming that non-indigenous languages are at equilibrium with respect to their interaction with indigenous languages (lengua indígena). We thus include one type "NIU (not in universe)" representing all non-indigenous languages, whose average fitness is controlled by one parameter.

The timeseries from Zambia derives from decennial national census 1990-2010 accessed through IPUMS International (Ruggles et al., 2003, Harmonized variable LANGZM1 ("Primary language") with Person weight). We excluded "Mashasha" which was not recorded in Census 1990 or 2000, and the non-language categories "Other", "Other African language", "Language of the Americas other than English", "Asian language", "European language",

Table 6: Mappings between language categories for languages occurring in Canadian census 2001-2016 used for time series inference.

Source name	Normalized name	Source name	Normalized name
Aboriginal languages, n.i.e.	EXCLUDE	Latvian	Latvian (Lettish)
Aboriginal languages, n.o.s.	EXCLUDE	Latvian (Lettish)	Latvian (Lettish)
African languages, n.i.e.	EXCLUDE	Lillooet	Salish s.l.
Afrikaans	EXCLUDE	Lingala	Bantu languages, n.i.e.
Afro-Asiatic languages, n.i.e.	EXCLUDE	Lithuanian	Lithuanian
Akan (Twi)	Twi	Macedonian	Macedonian
Albanian	EXCLUDE	Malagasy	Malayo-Polynesian languages, n.i.e.
Algonquian languages, n.i.e.	Algonquin	Malayalam	Malayalam
Algonquin	Algonquin	Malay-Bahasa	Malay-Bahasa
American Sign Language	EXCLUDE	Malay	Malay-Bahasa
Amharic	Amharic	Malayo-Polynesian languages, n.i.e.	Malayo-Polynesian languages, n.i.e.
Arabic	Arabic	Malecite	Malecite
Armenian	Armenian	Maltese	Maltese
Asiatic languages, n.i.e.	EXCLUDE	Mandarin	Mandarin
Assyrian Neo-Aramaic	EXCLUDE	Marathi	Marathi
Athabaskan languages, n.i.e.	Athapaskan s.l.	Michif	Cree
Athapaskan languages, n.i.e.	Athapaskan s.l.	Micmac	Micmac
Atikamekw	Attikamekw	Mi'kmaq	Micmac
Attikamekw	Attikamekw	Min Dong	EXCLUDE
Austro-Asiatic languages, n.i.e.	EXCLUDE	Min Nan (Chaochow, Teochow, Fukien, Taiwanese)	EXCLUDE
Austro-Asiatic languages, n.i.e	EXCLUDE	Mohawk	Mohawk
Austronesian languages, n.i.e.	EXCLUDE	Mongolian	EXCLUDE
Azerbaijani	Turkic languages, n.i.e.	Montagnais (Innu)	Montagnais-Naskapi
Babine (Wetsuwet'en)	Athapaskan s.l.	Montagnais-Naskapi	Montagnais-Naskapi
Bamanankan	Niger-Congo languages, n.i.e.	Moose Cree	Cree
Bantu languages, n.i.e.	Bantu languages, n.i.e.	Naskapi	Montagnais-Naskapi
Beaver	Athapaskan s.l.	Nepali	EXCLUDE
Belarusan	Byelorussian	Niger-Congo languages, n.i.e.	Niger-Congo languages, n.i.e.
Belarusian (Byelorussian)	Byelorussian	Niger-Congo languages, n.i.e.	Niger-Congo languages, n.i.e.
Bengali	Bengali	Nilo-Saharan languages, n.i.e.	EXCLUDE
Berber languages (Kabyle)	EXCLUDE	Nilo-Saharan languages, n.i.e.	EXCLUDE
Berber languages, n.i.e.	EXCLUDE	Nisga'a	Nishga
Bikol	Malayo-Polynesian languages, n.i.e.	Nishga	Nishga
Bilen	EXCLUDE	Nootka	Nootka
Bisayan languages	Malayo-Polynesian languages, n.i.e.	Nootka (Nuu-chah-nulth)	Nootka
Blackfoot	Blackfoot	Northern East Cree	Cree
Bosnian	Slavic languages, n.i.e.	Northern Tutchone	Athapaskan s.l.
Bulgarian	Bulgarian	North Slave (Hare)	Slavey
Burmese	EXCLUDE	North Slavey (Hare)	Slavey
Burmese	Tibeto-Burman	Norwegian	Norwegian
Byelorussian	Byelorussian	Nuu-chah-nulth (Nootka)	Nootka
Cantonese	Cantonese	Ojibway	Ojibway
Carrier	Carrier	Oji-Cree	Oji-Cree
Catalan	EXCLUDE	Okanagan	Salish s.l.
Cayuga	Iroquoian	Oneida	Iroquoian
Cebuano	Malayo-Polynesian languages, n.i.e.	Oriya (Odia)	Indo-Iranian languages, n.i.e.
Celtic languages, n.i.e.	Celtic	Oromo	EXCLUDE
Chaldean Neo-Aramaic	EXCLUDE	Other languages, n.i.e.	EXCLUDE
Chaochow (Teochow)	EXCLUDE	Ottawa (Odawa)	Algonquin
Chilcotin	Chilcotin	Pampangan (Kapampangan, Pampango)	Malayo-Polynesian languages, n.i.e.
Chinese languages, n.i.e.	EXCLUDE	Pampango	Malayo-Polynesian languages, n.i.e.
Chinese languages, n.o.s.	EXCLUDE	Pangasinan	Malayo-Polynesian languages, n.i.e.
Chinese, n.o.s. [2]	EXCLUDE	Panjabi (Punjabi)	Punjabi
Chinese, n.o.s. [3]	EXCLUDE	Pashto	Pashto

Source name	Normalized name	Source name	Normalized name
Chinese, n.o.s.	EXCLUDE	Persian (Farsi)	Persian (Farsi)
Chipewyan	Dene	Plains Cree	Cree
Comox	Salish s.l.	Polish	Polish
Cree	Cree	Portuguese	Portuguese
Cree, n.i.e.	Cree	Punjabi (Panjabi)	Punjabi
Cree, n.o.s.	Cree	Punjabi	Punjabi
Creole languages, n.i.e.	EXCLUDE	Quebec Sign Language	EXCLUDE
Creole languages, n.i.e.	EXCLUDE	Romance languages, n.i.e.	EXCLUDE
Creole, n.o.s.	EXCLUDE	Romanian	Romanian
Creoles	EXCLUDE	Rundi (Kirundi)	Bantu languages, n.i.e.
Croatian	Croatian	Russian	Russian
Cushitic languages, n.i.e.	EXCLUDE	Rwanda (Kinyarwanda)	Bantu languages, n.i.e.
Czech	Czech	Salish languages, n.i.e.	Salish s.l.
Dakota	Dakota/Sioux	Sarcee	Athapaskan s.l.
Dakota/Sioux	Dakota/Sioux	Sarsi (Sarcee)	Athapaskan s.l.
Danish	Danish	Scottish Gaelic	Gaelic languages
Dene	Dene	Sekani	Athapaskan s.l.
Dinka	EXCLUDE	Semitic languages, n.i.e.	Semitic languages, n.i.e.
Dogrib	Dogrib	Serbian	Serbian
Dogrib (Tlicho)	Dogrib	Serbo-Croatian	Serbo-Croatian
Dravidian languages, n.i.e.	Dravidian	Shanghainese	EXCLUDE
Dutch	Dutch	Shona	Bantu languages, n.i.e.
Edo	Niger-Congo languages, n.i.e.	Shuswap (Secwepemctsin)	Shuswap
English	English	Shuswap	Shuswap
Estonian	Estonian	Sign languages, n.i.e	EXCLUDE
Ewe	Niger-Congo languages, n.i.e.	Sindhi	Sindhi
Fijian	Malayo-Polynesian languages, n.i.e.	Sinhala (Sinhalese)	Sinhalese
Finnish	Finnish	Sinhalese	Sinhalese
Flemish	Flemish	Sino-Tibetan languages, n.i.e.	Tibeto-Burman
French	French	Siouan languages (Dakota/Sioux)	Dakota/Sioux
Frisian	Frisian	Siouan languages, n.i.e.	Dakota/Sioux
Fukien	EXCLUDE	Slavey, n.o.s.	Slavey
Fulah (Pular, Pulaar, Fulfulde)	EXCLUDE	Slavic languages, n.i.e.	Slavic languages, n.i.e.
Gaelic languages	Gaelic languages	Slovak	Slovak
Ganda	Bantu languages, n.i.e.	Slovene (Slovenian)	Slovenian
Ga	Niger-Congo languages, n.i.e.	Slovenian	Slovenian
Georgian	EXCLUDE	Somali	Somali
German	German	Southern East Cree	Cree
Germanic languages, n.i.e.	EXCLUDE	Southern Tutchone	Athapaskan s.l.
Gitksan	Gitksan	South Slave	Slavey
Gitxsan (Gitksan)	Gitksan	South Slavey	Slavey
Greek	Greek	Spanish	Spanish
Gujarati	Gujarati	Squamish	Salish s.l.
Gwich'in	Gwich'in	Stoney	Dakota/Sioux
Haida	Haida	Straits	Salish s.l.
Haisla	Wakashan s.l.	Swahili	Swahili
Haitian Creole	EXCLUDE	Swampy Cree	Cree
Hakka	Hakka	Swedish	Swedish
Halkomelem	Salish s.l.	Tagalog (Pilipino, Filipino)	Tagalog (Pilipino)
Harari	EXCLUDE	Tagalog (Pilipino)	Tagalog (Pilipino)
Hebrew	Hebrew	Tahltan	Athapaskan s.l.
Heiltsuk	Wakashan s.l.	Tai-Kadai languages, n.i.e	EXCLUDE
Hiligaynon	Malayo-Polynesian languages, n.i.e.	Taiwanese	EXCLUDE
Hindi	Hindi	Tamil	Tamil
Hmong-Mien languages	EXCLUDE	Telugu	Telugu
Hungarian	Hungarian	Thai	Thai
Icelandic	Icelandic	Thompson (Ntlakapamux)	Thompson
Igbo	Niger-Congo languages, n.i.e.	Tibetan languages	Tibeto-Burman
Ilocano	Malayo-Polynesian languages, n.i.e.	Tibetan	Tibeto-Burman

Table 6: (continued)

Source name	Normalized name	Source name	Normalized name
Indo-Iranian languages, n.i.e.	Indo-Iranian languages, n.i.e.	Tibeto-Burman languages, n.i.e.	Tibeto-Burman
Innu/Montagnais	Montagnais-Naskapi	Tigrigna	Tigringa
Inuinnaqtun	Inuit s.l.	Tigringa	Tigringa
Inuinnaqtun (Inuvialuktun)	Inuit s.l.	Tlicho (Dogrib)	Dogrib
Inuit languages, n.i.e.	Inuit s.l.	Tlingit	Tlingit
Inuktitut (Eskimo)	Inuit s.l.	Tsimshian	Tsimshian
Inuktitut	Inuit s.l.	Turkic languages, n.i.e.	Turkic languages, n.i.e.
Inuktitut, n.i.e.	Inuit s.l.	Turkish	Turkish
Inuvialuktun	Inuit s.l.	Tutchone, n.o.s.	Athapaskan s.l.
Iroquoian languages, n.i.e.	Iroquoian	Twi	Twi
Iroquoian s.l.	Iroquoian	Twi	Twi
Italian	Italian	type	normalized$_{name}$
Italic (romance) languages, n.i.e.	EXCLUDE	Ukrainian	Ukrainian
Japanese	Japanese	Uralic languages, n.i.e.	EXCLUDE
Kabyle	EXCLUDE	Urdu	Urdu
Kannada	Kannada	Uyghur	Turkic languages, n.i.e.
Karenic languages	Tibeto-Burman	Uzbek	Turkic languages, n.i.e.
Kashmiri	Indo-Iranian languages, n.i.e.	Vietnamese	Vietnamese
Kaska (Nahani)	Athapaskan s.l.	Vlaams (Flemish)	Flemish
Khmer (Cambodian)	Khmer (Cambodian)	Wakashan languages, n.i.e.	Wakashan s.l.
Kinyarwanda (Rwanda)	Bantu languages, n.i.e.	Waray-Waray	Malayo-Polynesian languages, n.i.e.
Konkani	Konkani	Welsh	Welsh
Korean	Korean	Wetsuweten	Athapaskan s.l.
Kurdish	Kurdish	Wolof	Niger-Congo languages, n.i.e.
Kutchin-Gwich'in (Loucheux)	Gwich'in	Woods Cree	Cree
Kutenai	Kutenai	Wu (Shanghainese)	EXCLUDE
Kwakiutl (Kwak'wala)	Wakashan s.l.	Yiddish	EXCLUDE
Lao	Lao	Yoruba	Niger-Congo languages, n.i.e.

Table 7: Mappings between language categories for languages occurring in United States American Communities Survey 2000-2015 used for time series inferrence.

Source name	Normalized name	Source name	Normalized name
.N/A (less than 5 years old/speaks only English)	English	.Pakistan N.E.C.	EXCLUDE
.N/A (less than 5 years old/speaks only English)	English	.India n.e.c.	EXCLUDE
.Navajo	Navajo	.India N.E.C.	EXCLUDE
.Navaho	Navajo	.Pakistan n.e.c.	EXCLUDE
.Miao-yao, Mien	Mien	.Chinese	EXCLUDE
.Mien	Mien	.African	EXCLUDE
.Miao, Hmong	Hmong	.Other Indo-European languages	EXCLUDE
.Hmong	Hmong	.Other Asian languages	EXCLUDE
.Hocano	Ilocano	.Other Pacific Island languages	EXCLUDE
.Ilocano	Ilocano	.Other specified African languages	EXCLUDE
.Sinlalese	Sinhalese	.Aleut-Eskimo languages	EXCLUDE
.Sinhalese	Sinhalese	.South/Central American Indian languages	EXCLUDE
.Lettish	Latvian	.Other Specified North American Indian languages	EXCLUDE
.Afrikaans	Afrikaans	.Other languages	EXCLUDE
.Afrikans	Afrikaans	.Uncodable	EXCLUDE
.Latvian	Latvian	.Other Algonquian languages	EXCLUDE
.Gujarati	Gujarathi	.Other Algonquin languages	EXCLUDE
.Gujarathi	Gujarathi	.Other specified American Indian languages	EXCLUDE
.Serbo-Croatian	Serbocroatian	.Other specified languages	EXCLUDE
.Serbocroatian	Serbocroatian	.South/Central America Indian	EXCLUDE
.American Indian	EXCLUDE	.Language not specified	EXCLUDE

and "Unknown".

For the United States of America, we use public use microdata samples from the American Communities Survey 1-year files from 2006 to 2015, available at `http://www2.census.gov/programs-surveys/acs/data/pums/`. We use the variable LANP to designate language, and we record the number of speakers as the sum of PWGTP (person weight) for each distinct language. LANP codes the response to the question "What is this language?" in the context "Does this person speak a language other than English at home?". The US census aggregates languages in the LANP variable so that they each achieve a high minimum frequency or censors languages that do not achieve that frequency. We discarded the minimum frequency bin in US data due to this explicit censorship bias. As with the Canadian census, we harmonized languages according to the mapping in Table 7.

CHAPTER 4 : Conclusion

We have seen that ideas from ecology and evolution can be powerful tools to understand social and cultural dynamics, provided we are able to look at systems in the right way. This investigation into cultural and social evolution has advanced science by providing a new way of looking at cultural change, which has both predictive power and rationalizes existing findings, and by advancing a general model of social evolution which has the capacity to broadly categorize social systems and can be adapted to useful inference.

All around us, culture is evolving by highly complex mechanisms, yet, like the information encoded in DNA, cultural information must be transmitted, and thereby exhibit some degree of stochastic drift, just as in evolutionary biology. We may well find, as a science, that only very few cultural traits support a reasonable analogy to biological alleles or biological mutation, yet it is important to study those that do, so that we may recognize the similarities and differences.

We see that exchangeable frequency-dependent selection covers a broad range of evolutionary phenomena, and allows us to categorize that broad range into a narrower range of possible behaviors. Chapter 1 presents a complicated evolutionary situation, involving sexual selection that depends on diploid genotypes and the complexities of Mendelian segregation, yet the situation nonetheless reduces to frequency-dependent selection because of a slow manifold. Thus, there is some scale at which the particulars of the model disappear and the behavior appears as a model of frequency-dependent selection. We observe a similar result in Chapter 3 when investigating the frequency-dependence induced by novelty bias: This resolves some of the debate about mechanisms of frequency-dependence (Brisson, 2018). The pure frequency-dependent model provides a notion of effective frequency-dependence, in much the same way that the population size in the Wright-Fisher process allows many population structures amount to the same effective population size. Many mechanisms may cause effective frequency-dependence, as observed by inference using the exchangeable

model. Many effects of frequency-dependence, by whatever mechanism, are broadly similar, for example, encouragement or suppression of diversity.

By further research into the connections between mechanisms of drift, selection, or frequency-dependent selection and their population-level consequences, we will understand more and more of the change around us.

APPENDIX

A.1. Detailed balance in the exchangeable model

Let X_t be the state of the population at time t, where $(X_t)_i \in \{0, ..., N\}, i \in \{1, ..., S\}$ is the number of agents adopting standard i. The transition probability matrix of the embedded Markov chain, $(X_{t+1} = (\omega_1, \omega_2, ..., \omega_n) | X_t = (\upsilon_1, \upsilon_2, ..., \upsilon_n))$, is zero unless $||X_{t+1} - X_t||_1 \in \{0, 2\}$ and $(X_{t+1} - X_t)\cdot = 0$. In other words, it is non-zero exactly when $X_{t+1} - X_t$ is the sum of two vectors, e_i and $-e_j$ where the es are the basis vectors $e_1 = (1, 0, 0, ...), e_2 = (0, 1, 0, ...), ..., e_S = (..., 0, 1)$. At any starting point, the non-zero probabilities may be indexed by i, j where $P_{ij}(x_t) = (X_{t+1} = x_t - e_i + e_j | X_t = x_t)$. This can be thought of as a matrix

$$P_{ij}(x_t) = (i\text{-type removed})(j\text{-type added})$$
$$= \frac{(x_t)_i}{N} \left((1 - \mu) \frac{w((x_t)_j)(x_t)_j}{\sum_{k=1}^{S} w((x_t)_k)(x_t)_k} + \mu \frac{1}{S} \right)$$

This looks like one step in a random walk on an S-dimensional simplex, where the one-step probabilities change slightly depending on the location. We would like to to establish that detailed balance holds, and observe the equilibrium distribution. First note that $(X_{t+1} = i | X_t = j) = 0$ iff $(X_{t+1} = j | X_t = i) = 0$, and thus, detailed balance holds for all such transitions. In light of this, the detailed balance condition can be rewritten:

$$P_{ij}(x)\pi(x) = P_{ji}(x - e_i + e_j)\pi(x - e_i + e_j),$$

where $\pi(x)$ is the putative equilibrium probability of observing the population in a state x. To establish this condition, it suffices to show that for any x, $P_{ij}(x)P_{jk}(x - e_i + e_j)P_{ki}(x - e_i + e_k) = P_{ik}(x)P_{kj}(x - e_i + e_k)P_{ji}(x - e_i + e_j)$. This comes from the Kolmogorov criterion and the notion that larger loops fulfilling the Kolmogorov criterion may be built out of this

set of small loops that fulfill the criterion. By simplification, the first loop's product is:

$$\frac{x_i}{N}\left(\mu^c\frac{w(x_j)x_j}{Z(x)}+\mu\frac{1}{S}\right)\frac{x_j+1}{N}\left(\mu^c\frac{w(x_k)x_k}{Z(x-e_i+e_j)}+\mu\frac{1}{S}\right)$$
$$\frac{x_k+1}{N}\left(\mu^c\frac{w(x_i-1)(x_i-1)}{Z(x-e_i+e_k)}+\mu\frac{1}{S}\right),$$

where the partition function $Z(x) = \sum_{l=1}^{S} w(x_l)x_l$. The second loop's product is:

$$\frac{x_i}{N}\left(\mu^c\frac{w(x_k)x_k}{Z(x)}+\mu\frac{1}{S}\right)\frac{x_k+1}{N}\left(\mu^c\frac{w(x_j)x_j}{Z(x-e_i+e_k)}+\mu\frac{1}{S}\right)$$
$$\frac{x_j+1}{N}\left(\mu^c\frac{w(x_i-1)(x_i-1)}{Z(x-e_i+e_j)}+\mu\frac{1}{S}\right).$$

In general, the condition for detailed balance is, taking $w(x_i-1)(x_i-1)$ to be f_i, $w(x_j)x_j$, $w(x_k)x_k$ to be f_j, f_k respectively, $Z(x-e_i+e_l)$ to be z_l, $\mu/(\mu^c S) = c$,

$$\left(\frac{f_i}{z_k}+c\right)\left(\frac{f_j}{z_i}+c\right)\left(\frac{f_k}{z_j}+c\right)=\left(\frac{f_i}{z_j}+c\right)\left(\frac{f_j}{z_k}+c\right)\left(\frac{f_k}{z_i}+c\right)$$

Thus, the existence of detailed balance depends only on the choice of w. The condition is fulfilled trivially when w is constant, for then $z_i = z_j = z_k$. If $w(x)$ is non-constant, then the partition function $Z(x) = \sum_{k=1}^{S} w(x_k)x_k$ is variable independently of any $w(x_k)$.

BIBLIOGRAPHY

D. M. Abrams and S. H. Strogatz. Linguistics: Modelling the dynamics of language death. *Nature*, 424(6951):900–900, 2003.

P. M. Altrock, A. Traulsen, and F. A. Reed. Stability properties of underdominance in finite subdivided populations. *PLoS Comput Biol*, 7(11):e1002260–e1002260, 2011.

C. Anderson, S. Wong, A. Fuller, K. Zigelsky, and R. Earley. Carotenoid-based coloration is associated with predation risk, competition, and breeding status in female convict cichlids (amatitlania siquia) under field conditions. *Environmental Biology of Fishes*, 98 (4):1005–1013, 2015.

M. Andersson and L. W. Simmons. Sexual selection and mate choice. *Trends in Ecology & Evolution*, 21(6):296–302, 2006.

M. B. Andersson. *Sexual selection*. Princeton University Press, 1994.

C. F. Arias, A. G. Munoz, C. D. Jiggins, J. Mavarez, E. Bermingham, and M. Linares. A hybrid zone provides evidence for incipient ecological speciation in heliconius butterflies. *Molecular ecology*, 17(21):4699–4712, 2008.

M. E. Arnegard and A. S. Kondrashov. Sympatric speciation by sexual selection alone is unlikely. *Evolution*, 58(2):222–237, 2004.

M. Aronoff. Darwinism tested by the science of. *On looking into words (and beyond): Structures, Relations, Analyses*, page 443, 2017.

R. Axelrod et al. The evolution of strategies in the iterated prisoner's dilemma. *The dynamics of norms*, pages 1–16, 1987.

V. Bansal, A. Bashir, and V. Bafna. Evidence for large inversion polymorphisms in the human genome from hapmap data. *Genome research*, 17(2):219–230, 2007.

N. Barton and S. Rouhani. The probability of fixation of a new karyotype in a continuous population. *Evolution*, pages 499–517, 1991.

G. J. Baxter, R. A. Blythe, W. Croft, and A. J. McKane. Utterance selection model of language change. *Physical Review E*, 73(4):046118, 2006.

G. J. Baxter, R. A. Blythe, W. Croft, and A. J. McKane. Modeling language change: an evalution of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, 21(2):257–296, 2009.

R. A. Bentley. Random drift versus selection in academic vocabulary: An evolutionary analysis of published keywords. *PLoS ONE*, 3(8):e3057, 2008.

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

R. A. Bentley, P. Ormerod, and S. Shennan. Population-level neutral model already explains linguistic patterns. *Proceedings of the Royal Society of London B: Biological Sciences*, 278 (1713):1770–1772, 2011.

J. Berger, E. T. Bradlow, A. Braunstein, and Y. Zhang. From karen to katie: Using baby names to understand cultural evolution. *Psychological science*, 23(10):1067–1073, 2012.

J. D. Blount. Carotenoids and life-history evolution in animals. *Archives of Biochemistry and Biophysics*, 430(1):10–15, 2004.

J. D. Blount, N. B. Metcalfe, T. R. Birkhead, and P. F. Surai. Carotenoid modulation of immune function and sexual attractiveness in zebra finches. *Science*, 300(5616):125–127, 2003.

R. A. Blythe. Neutral evolution: a null model for language dynamics. *Advances in Complex Systems*, 15:1150015, 2012.

R. A. Blythe and W. Croft. S-curves and the mechanisms of propagation in language change. *Language*, 88(2):269–304, 2012.

D. Brisson. Negative frequency-dependent selection is frequently confounding. *Frontiers in Ecology and Evolution*, 6:10, 2018.

L. W. Buss. *The evolution of individuality*. Princeton University Press, 2014.

J. L. Bybee. *Morphology: A study of the relation between meaning and form*, volume 9. John Benjamins, 1985.

J. L. Bybee and C. L. Moder. Morphological classes as natural categories. *Language*, pages 251–270, 1983.

A. Caballero and W. G. Hill. Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics*, 131(2):493–507, 1992.

C. Cannings. The latent roots of certain markov chains arising in genetics: a new approach, i. haploid models. *Advances in Applied Probability*, 6(2):260–290, 1974.

E. Canonge and E. Pittman, R.S. *Ethnologue of Bibleless tribes: for prayer intercessors, Bible translators, missionaries, prospective missionaries, mission councils*. Wycliffe Bible Translators, 1958.

L. L. Cavalli-Sforza and M. W. Feldman. *Cultural transmission and evolution: a quantitative approach*. Number 16. Princeton University Press, 1981.

B. Charlesworth. Evolutionary rates in partially self-fertilizing species. *American Naturalist*, pages 126–148, 1992.

M. H. Christiansen, N. Chater, and P. W. Culicover. *Creating language: Integrating evolution, acquisition, and processing.* MIT Press, 2016.

T. Clutton-Brock and K. McAuliffe. Female mate choice in mammals. *The Quarterly Review of Biology*, 84(1):3–27, 2009.

B. Coulmont, V. Supervie, and R. Breban. The diffusion dynamics of choice: From durable goods markets to fashion first names. *Complexity*, 2015.

V. P. Crawford and J. Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 50(6):1431–1451, 1982.

N. Creanza, M. Ruhlen, T. J. Pemberton, N. A. Rosenberg, M. W. Feldman, and S. Ramachandran. A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences*, 112(5):1265–1272, 2015.

B. J. Crespi. Causes of assortative mating in arthropods. *Animal Behaviour*, 38(6):980–1000, 1989.

F. Crick, L. Barnett, S. Brenner, and R. Watts-tobin. General nature of the genetic code for proteins. *Nature*, 192:1227–1232, 1961.

F. H. Crick. The origin of the genetic code. *Journal of molecular biology*, 38(3):367–379, 1968.

W. Croft. *Explaining language change: An evolutionary approach.* Pearson Education, 2000.

J. F. Crow, M. Kimura, et al. An introduction to population genetics theory. *An introduction to population genetics theory.*, 1970.

T. Crowley. Melanesian languages: Do they have a future? *Oceanic Linguistics*, pages 327–344, 1995.

C. F. Cuskley, M. Pugliese, C. Castellano, F. Colaiori, V. Loreto, and F. Tria. Internal and external dynamics in language: evidence from verb regularity in a historical corpus of english. *PLoS ONE*, 9(8):e102882, 2014.

O. Dahl. *Inflationary effects in language and elsewhere.* John Benjamins, Philadelphia, 2001.

C. Damgaard. Fixation of advantageous alleles in partially self-fertilizing populations: the effect of different selection modes. *Genetics*, 154(2):813–821, 2000.

C. Darwin. *The descent of man and selection in relation to sex*, volume 1. Murray, 1888.

M. Davies. Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora*, 7(2):121–157, 2012.

R. Dawkins. *The Selfish Gene*. Oxford University Press, 1976.

R. Der, C. L. Epstein, and J. B. Plotkin. Generalized population models and the nature of genetic drift. *Theoretical Population Biology*, 80(2):80–99, 2011.

R. Der, C. Epstein, and J. B. Plotkin. Dynamics of neutral and selected alleles when the offspring distribution is skewed. *Genetics*, 191(4):1331–1344, 2012.

P. Diaconis and D. Freedman. Finite exchangeable sequences. *The Annals of Probability*, pages 745–764, 1980.

R. M. Dixon. The endangered languages of australia, indonesia and oceania. In R. H. Robins and E. M. Uhlenbeck, editors, *Endangered languages*, chapter 9, pages 229–255. Berg Oxford, 1991.

C. Dunford. Kin selection for ground squirrel alarm calls. *The American Naturalist*, 111 (980):782–785, 1977.

B. Efron and D. V. Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, 65(3):457–483, 1978.

B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pages 54–75, 1986.

A. Ellegård. *The auxiliary do: The establishment and regulation of its use in English*. Almquist & Wiksell, Stockholm, 1953.

C. L. Epstein and R. Mazzeo. *Degenerate diffusion operators arising in population biology*. Princeton University Press, 2013.

W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoretical population biology*, 3(1):87–112, 1972.

W. J. Ewens. Mathematical population genetics. I. theoretical introduction., 2004.

W. J. Ewens et al. *Population genetics*. Springer, 1969.

A. F. Feder, S. Kryazhimskiy, and J. B. Plotkin. Identifying signatures of selection in genetic time series. *Genetics*, 196(2):509–522, 2014.

V. Ferdinand and W. Zuidema. Thomas' theorem meets bayes' rule: A model of the iterated learning of language. In *Proceedings of the Cognitive Science Society*, volume 31, 2009.

V. Ferdinand, B. Thompson, S. Kirby, and K. Smith. Regularization behavior in a non-

linguistic domain. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35, 2013.

L. Feuk, J. R. MacDonald, T. Tang, A. R. Carson, M. Li, G. Rao, R. Khaja, and S. W. Scherer. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee dna sequence assemblies. *PLoS genetics*, 1(4):e56, 2005.

J. A. Fishman. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*, volume 76. Multilingual matters, 1991.

J. F. Fontanari and L. I. Perlovsky. Solvable null model for the distribution of word frequencies. *Physical Review E*, 70(4):042901, 2004.

S. J. Freeland and L. D. Hurst. The genetic code is one in a million. *Journal of molecular evolution*, 47(3):238–248, 1998.

D. J. Futuyma. *Evolution Second Edition*. Sinauer, 2009.

C. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences Springer Series in Synergetics*. Springer, 2009.

S. Gavrilets. *Fitness landscapes and the origin of species (MPB-41)*. Princeton University Press, 2004.

S. Ghirlanda, A. Acerbi, H. Herzog, and J. A. Serpell. Fashion vs. function in cultural evolution: The case of dog breed popularity. *PLoS One*, 8(9):e74770, 2013.

S. Glémin. Extinction and fixation times with dominance and inbreeding. *Theoretical population biology*, 81(4):310–316, 2012.

C. Goodwillie, S. Kalisz, and C. G. Eckert. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, pages 47–79, 2005.

A. Grafen. Natural selection, kin selection and group selection. *Behavioural ecology: An evolutionary approach*, 2:62–84, 1984.

V. Grant. *The evolutionary process: a critical study of evolutionary theory*. Columbia University Press, 1991.

R. D. Gray and Q. D. Atkinson. Language-tree divergence times support the anatolian theory of indo-european origin. *Nature*, 426(6965):435–439, 2003.

D. Greig and M. Travisano. The prisoner's dilemma and polymorphism in yeast suc genes. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(Suppl 3):S25–S26, 2004.

R. Griffiths. On the distribution of allele frequencies in a diffusion model. *Theoretical Population Biology*, 15(1):140–158, 1979.

E. Haeckel. *Natürliche Schöpfungsgeschichte*. Gg. Reimer, 1868.

M. W. Hahn and R. A. Bentley. Drift as a mechanism for cultural change: an example from baby names. *Proceedings of the Royal Society B*, 270(Suppl 1):S120–S123, 2003.

J. Haldane. A mathematical theory of natural and artificial selection. part ii the influence of partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of mendelian populations, and on natural selection. *Biological Reviews*, 1(3):158–163, 1924.

J. A. Hawkins. A parsing theory of word order universals. *Linguistic Inquiry*, 21(2):223–261, 1990.

J. Hay and H. Baayen. Parsing and productivity. In *Yearbook of morphology 2001*, pages 203–235. Springer, 2002.

P. W. Hedrick. The establishment of chromosomal variants. *Evolution*, pages 322–332, 1981.

M. Higashi, G. Takimoto, and N. Yamamura. Sympatric speciation by sexual selection. *Nature*, 402(6761):523–526, 1999.

J. B. Hooper. Word frequency in lexical diffusion and the source of morphophonological change. *Current progress in historical linguistics*, 96:105, 1976.

M. Hori. Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. *Science*, 260(5105):216–219, 1993.

S. P. Hubbell. *The unified neutral theory of biodiversity and biogeography (MPB-32)*, volume 32. Princeton University Press, 2001.

Y. Itan, A. Powell, M. A. Beaumont, J. Burger, and M. G. Thomas. The origins of lactase persistence in europe. *PLoS computational biology*, 5(8):e1000491, 2009.

R. Jakobson, L. R. Waugh, and M. Monville-Burston. *On Language*. Harvard University Press, Cambridge, MA, 1995.

A. C. Janetos. Strategies of female mate choice: a theoretical analysis. *Behavioral Ecology and Sociobiology*, 7(2):107–112, 1980.

D. H. Janzen. Why bamboos wait so long to flower. *Annual Review of Ecology and Systematics*, 7(1):347–391, 1976.

O. Jespersen. *Negation in English and other languages*. København, AF Høst, 1917.

C. D. Jiggins, W. O. McMillan, W. Neukirchen, and J. Mallet. What can hybrid zones tell us about speciation? the case of heliconius erato and h. himera (lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*, 59(3):221–242, 1996.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Discrete multivariate distributions*. New York: John Wiley & Sons, 1997.

A. G. Jones and N. L. Ratterman. Mate choice and sexual selection: what have we learned since darwin? *Proceedings of the National Academy of Sciences*, 106(Supplement 1): 10001–10008, 2009.

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

A. Kandler, R. Unger, and J. Steele. Language shift, bilingualism and the future of britain's celtic languages. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1559):3855–3864, 2010.

D. D. Kapan. Three-butterfly system provides a field test of müllerian mimicry. *Nature*, 409(6818):338–340, 2001.

S. Karlin. Comparisons of positive assortative mating and sexual selection models. *Theoretical population biology*, 14(2):281–312, 1978.

S. Karlin and S. Lessard. *Theoretical studies on sex ratio evolution*, volume 22. Princeton University Press, 1986.

S. Karlin and J. McGregor. The number of mutant forms maintained in a population. In *Proceedings of the Fifth Berkeley Symposium on mathematics, Statistics and probability*, volume 4, pages 415–438, 1967.

S. Karlin and J. McGregor. Addendum to a paper of w. ewens. *Theoretical population biology*, 3(1):113–116, 1972.

M. L. Katz and C. Shapiro. Network externalities, competition, and compatibility. *The American economic review*, pages 424–440, 1985.

M. L. Katz and C. Shapiro. Technology adoption in the presence of network externalities. *The journal of political economy*, pages 822–841, 1986.

J. P. Kenney-Hunt, T. T. Vaughn, L. S. Pletscher, A. Peripato, E. Routman, K. Cothran, D. Durand, E. Norgard, C. Perel, and J. M. Cheverud. Quantitative trait loci for body size components in mice. *Mammalian genome*, 17(6):526–537, 2006.

B. Kerr, M. A. Riley, M. W. Feldman, and B. J. Bohannan. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature*, 418(6894):171–174, 2002.

M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47 (6):713, 1962.

M. Kimura. Diffusion models in population genetics. *Journal of Applied Probability*, 1(2): 177–232, 1964.

M. Kimura. Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. *Genetics*, 52(5):875, 1965.

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

M. King. *Species evolution: the role of chromosome change*. Cambridge University Press, 1995.

M. Kirkpatrick. Sexual selection and the evolution of female choice. *Evolution*, pages 1–12, 1982.

M. Kirkpatrick and N. Barton. Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1):419–434, 2006.

R. D. Knight, S. J. Freeland, and L. F. Landweber. Rewiring the keyboard: evolvability of the genetic code. *Nature Reviews Genetics*, 2(1):49–58, 2001.

H. Kokko, M. D. Jennions, and R. Brooks. Unifying and testing models of sexual selection. *Annual Review of Ecology, Evolution, and Systematics*, pages 43–66, 2006.

M. Krauss. The world's languages in crisis. *Language*, 68(1):4–10, 1992.

M. E. Krauss. Native languages of alaska. In O. Miyaoka, O. Sakiyama, and M. E. Krauss, editors, *The vanishing languages of the Pacific Rim*, chapter 21, page 406. Oxford University Press, 2007.

A. S. Kroch. Reflexes of grammar in patterns of language change. *Language variation and change*, 1(03):199–244, 1989.

B. Kuijper, I. Pen, and F. J. Weissing. A guide to sexual selection theory. *Annual Review of Ecology, Evolution, and Systematics*, 43:287–311, 2012.

W. Labov. *Principles of linguistic change Volume 2: Social factors*. Blackwell, 2001.

W. Labov. *Principles of linguistic change Volume 3: Cognitive and cultural factors*. Blackwell, 2010.

R. Lande. Effective deme sizes during long-term evolution estimated from rates of chromosome rearrangement. *Evolution*, 33:234–251, 1979.

R. Lande. Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, 78(6):3721–3725, 1981.

R. Lande and D. W. Schemske. The evolution of self-fertilization and inbreeding depression in plants. i. genetic models. *Evolution*, pages 24–40, 1985.

K. Lange. *Applied probability*. Springer Science & Business Media, 2010.

K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of computational and graphical statistics*, 9(1):1–20, 2000.

G. M. Langham and C. Benkman. Specialized avian predators repeatedly attack novel color morphs of heliconius butterflies. *Evolution*, 58(12):2783–2787, 2004.

T. Lewens. *Cultural evolution: conceptual challenges*. OUP Oxford, 2015.

E. Lieberman, J.-B. Michel, J. Jackson, T. Tang, and M. A. Nowak. Quantifying the evolutionary dynamics of language. *Nature*, 449(7163):713–716, 2007.

G. Lupyan and R. Dale. The role of adaptation in understanding linguistic diversity. *Language structure and environment: Social, cultural, and natural factors*, pages 287–16, 2015.

S. J. Lycett. Acheulean variation and selection: does handaxe symmetry fit neutral expectations? *Journal of Archaeological Science*, 35(9):2640–2648, 2008.

S. J. Lycett. Understanding ancient hominin dispersals using artefactual data: a phylogeographic analysis of acheulean handaxes. *PLoS One*, 4(10):e7404, 2009.

J. Mallet and N. H. Barton. Strong natural selection in a warning-color hybrid zone. *Evolution*, pages 421–431, 1989.

A. Mesoudi. *Cultural evolution: How Darwinian theory can explain human culture and synthesize the social sciences*. University of Chicago Press, 2011.

A. Mesoudi and S. J. Lycett. Random copying, frequency-dependent copying and culture change. *Evolution and Human Behavior*, 30(1):41–48, 2009.

L. K. M'Gonigle and R. G. FitzJohn. Assortative mating and spatial structure in hybrid zones. *Evolution*, 64(2):444–455, 2010.

Y. Michalakis and I. Olivieri. The influence of local extinctions on the probability of fixation of chromosomal rearrangements. *Journal of evolutionary biology*, 6(2):153–170, 1993.

J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182, 2011.

O. Miyaoka, O. Sakiyama, and M. E. Krauss, editors. *The vanishing languages of the Pacific rim*. Oxford University Press, Oxford New York, 2007. ISBN 0-19-926662-X.

C. Moseley, editor. *Atlas of the World's Languages in Danger*. UNESCO Publishing, Paris, 3rd edition, 2010.

S. S. Mufwene. *The ecology of language evolution*. Cambridge University Press, 2001.

S. S. Mufwene. Colonization, globalization, and the future of languages in the twenty-first century. *International Journal on Multicultural Societies*, 4(2):162–193, 2002.

M. Müller. Darwinism tested by the science of language translated from the German of professor August Schleicher. *Nature*, 1:256–259, 1870.

E. Murgola. trna, suppression, and the code. *Annual review of genetics*, 19(1):57–80, 1985.

T. Nagylaki. *Introduction to theoretical population genetics*, volume 142. Springer-Verlag Berlin, 1992.

J. R. Nahum, B. N. Harding, and B. Kerr. Evolution of restraint in a structured rock–paper–scissors community. *Proceedings of the National Academy of Sciences*, 108(Supplement 2):10831–10838, 2011.

A. Navarro and N. H. Barton. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, 57(3):447–459, 2003a.

A. Navarro and N. H. Barton. Chromosomal speciation and molecular divergence–accelerated evolution in rearranged chromosomes. *Science*, 300(5617):321–324, 2003b.

F. D. Neiman. Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in illinois woodland ceramic assemblages. *American Antiquity*, pages 7–36, 1995.

D. Nettle. Is the rate of linguistic change constant? *Lingua*, 108(2):119–136, 1999.

D. Nettle. *Vanishing voices : the extinction of the world's languages*. Oxford University Press, New York, 2000. ISBN 0195136241.

E. Nevo, E. N. Ivanitskaia, and A. Beiles. *Adaptive radiation of blind subterranean mole rats: naming and revisiting the four sibling species of the Spalax ehrenbergi superspecies in Israel: Spalax galili (2n)*. Backhuys publishers, 2001.

M. G. Newberry, D. M. McCandlish, and J. B. Plotkin. Assortative mating can impede or facilitate fixation of underdominant alleles. *Theoretical Population Biology*, 112:14–21, 2016.

M. G. Newberry, C. A. Ahern, R. Clark, and J. B. Plotkin. Detecting evolutionary forces in language change. *Nature*, 551(7679):223, 2017.

M. A. Noor, K. L. Grams, L. A. Bertucci, and J. Reiland. Chromosomal inversions and the

reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21):12084–12088, 2001.

E. M. North. *The Book of A Thousand Tongues.* American Bible Society, Harper & Brothers, 1938.

M. A. Nowak and K. Sigmund. Evolutionary dynamics of biological games. *science*, 303 (5659):793–799, 2004.

P. O'Donald. *Genetic models of sexual selection*, volume 44. Cambridge University Press Cambridge, 1980.

J. P. O'Dwyer and A. Kandler. Novelty, popularity, and emergent neutrality: bias in the choice of baby names and lessons for analyzing cultural data. *arXiv preprint arXiv:1702.08506*, 2017.

J. M. Ortman and H. B. Shin. Language projections: 2010 to 2020. In *American Sociological Association Annual Meeting*, 2011.

S. P. Otto, M. R. Servedio, and S. L. Nuismer. Frequency-dependent selection and the evolution of assortative mating. *Genetics*, 179(4):2091–2112, 2008.

M. Pagel. Human language as a culturally transmitted replicator. *Nature Reviews Genetics*, 10(6):405–415, 2009.

M. Pagel. Linguistics and the evolution of human languages. In J. Losos, editor, *The Princeton Guide to Evolution*, chapter Chapter VIII.9, pages 786–794. Princeton University Press, Princeton, 2013.

M. Pagel, Q. D. Atkinson, and A. Meade. Frequency of word-use predicts rates of lexical evolution throughout indo-european history. *Nature*, 449(7163):717–720, 2007.

T. L. Parsons and T. Rogers. Dimension reduction via timescale separation in stochastic dynamical systems. *arXiv preprint arXiv:1510.07031*, 2015.

S. Pinker. Rules of language. *Science*, 253(5019):530–535, 1991.

D. N. Politis. The impact of bootstrap methods on time series analysis. *Statistical Science*, pages 219–230, 2003.

P. Pollett. On a model for interference between searching insect parasites. *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics*, 32(02):133–150, 1990.

A. Pomiankowski. The costs of choice in sexual selection. *Journal of theoretical Biology*, 128(2):195–218, 1987.

S. Prasada and S. Pinker. Generalisation of regular and irregular morphological patterns. *Language and cognitive processes*, 8(1):1–56, 1993.

L. Real. Search theory and mate choice. i. models of single-sex discrimination. *American Naturalist*, pages 376–405, 1990.

F. Reali and T. L. Griffiths. Evolution of frequency distributions: Relating regularization to inductive biases through iterate learning. *Cognition*, 111:317–328, 2009.

F. Reali and T. L. Griffiths. Words as alleles: connecting language evolution with bayesian learners to models of genetic drift. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1680):429–436, 2010.

R. G. Reeves, J. Bryk, P. M. Altrock, J. A. Denton, and F. A. Reed. First steps towards underdominant genetic transformation of insect populations. *PLoS ONE*, 9(5):e97557, 2014.

L. H. Rieseberg. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16(7):351–358, 2001.

D. Ringe, T. Warnow, and A. Taylor. Indo-european and computational cladistics. *Transactions of the Philological Society*, 100(1):59–129, 2002.

D. Roze and F. Rousset. Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics*, 165(4):2153–2166, 2003.

S. Ruggles, M. L. King, D. Levison, R. McCaa, and M. Sobek. Ipums-international. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 36(2):60–65, 2003.

T. Sasaki and D. Biro. Cumulative culture can emerge from collective intelligence in animal groups. *Nature communications*, 8:15049, 2017.

A. Schleicher. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Hermann Böhlau, 1861.

A. Schleicher. *Darwinism tested by the science of language, translated by Alex V. W. Bikkers*. John Camden Hotten, 1869.

D. W. Schultz and M. Yarus. Transfer rna mutation and the malleability of the genetic code. *Journal of molecular biology*, 235(5):1377–1380, 1994.

D. W. Schultz and M. Yarus. On malleability in the genetic code. *Journal of molecular evolution*, 42(5):597–601, 1996.

K. M. Sefc, A. C. Brown, and E. D. Clotfelter. Carotenoid-based coloration in cichlid fishes.

Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 173:42–51, 2014.

J. Seger. Unifying genetic models for the evolution of female choice. *Evolution*, pages 1185–1193, 1985.

M. R. Servedio and R. Bürger. The counterintuitive role of sexual selection in species maintenance and speciation. *Proceedings of the National Academy of Sciences*, 111(22): 8113–8118, 2014.

S. P. Sinkins and F. Gould. Gene drive systems for insect disease vectors. *Nature Reviews Genetics*, 7(6):427–435, 2006.

M. Slatkin. Fixation probabilities and fixation times in a subdivided population. *Evolution*, pages 477–488, 1981.

L. M. Smith, K. Bomblies, and D. Weigel. Complex evolutionary events at a tandem cluster of arabidopsis thaliana genes resulting in a single-locus genetic incompatibility. *PLoS genetics*, 7(7):e1002164, 2011.

H. Stefansson, A. Helgason, G. Thorleifsson, V. Steinthorsdottir, G. Masson, J. Barnard, A. Baker, A. Jonasdottir, A. Ingason, V. G. Gudnadottir, et al. A common inversion under selection in europeans. *Nature genetics*, 37(2):129–137, 2005.

A. J. Stewart, R. M. Seymour, A. Pomiankowski, and M. Reuter. Under-dominance constrains the evolution of negative autoregulation in diploids. *PLoS Comput Biol*, 9(3): e1002992, 2013.

M. Tamariz, T. M. Ellison, D. J. Barr, and N. Fay. Cultural selection drives the evolution of human communication systems. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1788):20140488, 2014.

G. Tassey. Standardization in technology-based markets. *Research policy*, 29(4):587–602, 2000.

M. Tomasello, M. Davis-Dasilva, L. Camak, and K. Bard. Observational learning of tool-use by young chimpanzees. *Human evolution*, 2(2):175–183, 1987.

P. Trudgill. *New-dialect formation: The inevitability of colonial Englishes*. Oxford University Press, USA, 2004.

Truth and Reconciliation Commission of Canada. *Honouring the Truth, Reconciling for the Future: Summary of the Final Report of the Truth and Reconciliation Commission of Canada*. Truth and Reconciliation Commission of Canada, www.trc.ca, Winnipeg, 2015. ISBN 978-0-660-02078-5.

M. T. Ullman. Acceptability ratings of regular and irregular past-tense forms: Evidence for

a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes*, 14(1):47–67, 1999.

S. Valverde and R. V. Solé. A cultural diffusion model for the rise and fall of programming languages. 2015.

J. H. Van Den Brink. *The Haida Indians: Cultural Change Mainly Between 1876-1970. With 5 Maps*, volume 8 of *Monographs and theoretical studies in sociology and anthropology in honour of Nels Anderson*. Leiden : E. J. Brill, 1974.

I. Volkov, J. R. Banavar, S. P. Hubbell, and A. Maritan. Neutral theory and relative species abundance in ecology. *Nature*, 424(6952):1035–1037, 2003.

I. Volkov, J. R. Banavar, F. He, S. P. Hubbell, and A. Maritan. Density dependence explains tree species abundance and diversity in tropical forests. *Nature*, 438(7068):658, 2005.

C. Waddington. Canalization of development and the inheritance of acquired characters. *Nature*, 150(3811):563–565, 1942.

A. Wagner, S. Ortman, and R. Maxfield. From the primordial soup to self-driving cars: standards and their role in natural and technological innovation. *Journal of The Royal Society Interface*, 13(115):20151086, 2016.

J. Walsh. Rate of accumulation of reproductive isolation by chromosome rearrangements. *American Naturalist*, 120(4):510–532, 1982.

G. Watterson. On the number of segregating sites in genetical models without recombination. *Theoretical population biology*, 7(2):256–276, 1975.

G. Watterson and H. Guess. Is the most frequent allele the oldest? *Theoretical population biology*, 11(2):141–160, 1977.

G. A. Watterson. The sampling theory of selectively neutral alleles. *Advances in applied probability*, pages 463–488, 1974.

S. A. West, A. S. Griffin, and A. Gardner. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of evolutionary biology*, 20(2): 415–432, 2007.

M. J. D. White. *Modes of speciation*. WH Freeman San Francisco, 1978.

M. C. Whitlock. Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. *Evolution*, 54(6):1855–1861, 2000.

M. C. Whitlock. Fixation probability and time in subdivided populations. *Genetics*, 164 (2):767–779, 2003.

D. D. Wiegmann, S. M. Seubert, and G. A. Wade. Mate choice and optimal search behavior: fitness returns under the fixed sample and sequential search strategies. *Journal of theoretical biology*, 262(4):596–600, 2010.

S. Wright. The evolution of dominance. *The American Naturalist*, 63(689):556–561, 1929.

S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97–159, 1931.

S. Wright. On the probability of fixation of reciprocal translocations. *American Naturalist*, pages 513–522, 1941.

S. Wright. *Evolution and the Genetics of Populations*, volume II. University of Chicago Press, 1969.

G. K. Zipf. *Human behavior and the principle of least effort: An introduction to human ecology.* Addison-Wesley Press, Cambridge, Mass., 1949.