Title
Morphology Gets More and More Complex, Unless It Doesn't

Permalink
https://escholarship.org/uc/item/8231h9n8

ISBN
978-1-108-47989-9

Authors
Meinhardt, Eric
Malouf, Robert
Ackerman, Farrell

Publication Date
2022

Peer reviewed

8 Morphology Gets More and More Complex, Unless It Doesn't

Eric Meinhardt, Robert Malouf, and Farrell Ackerman

The past few years have led to the widespread recognition that morphology is an independent domain of language functioning in dynamic interdependence with more familiar domains such as phonology and syntax. This has permitted nuanced research into the organization of morphological systems as well as the development of hypotheses concerning factors responsible for such organization. In this chapter we compare two classes of hypotheses – adaptive explanations and neutral ones – for attested differences in morphological complexity claimed to correspond with sociocultural and demographic factors. While both examine language change as a (cultural) evolutionary process, we argue that much recent work on adaptive hypotheses for morphological complexity has been uncritically adaptationist, neglecting key results and lessons from population genetics about how to study evolutionary systems. Finally, we argue that neutral explanations are presently more likely explanations for the apparent association of morphological complexity and smaller, historically more isolated populations and should a priori be preferred over adaptive explanations unless and until a high evidential burden has been met.

1 Introduction

Consider the following two large questions that have become central to discussions in the morphological literature and whose answers are foundational for theory construction in this domain:

(1) a. What do analysts mean when they talk about morphological complexity and make claims about learnability?
  b. What kinds of explanations do analysts advance given what they believe to be true about complexity and social conditions?

Recent work in morphology (Ackerman and Malouf 2013, Stump and Finkel 2013) has conceptualized morphological complexity in terms of two interdependent dimensions: *enumerative complexity* and *integrative complexity*, henceforth abbreviated as *E-complexity* and *I-complexity*. The E-complexity of a language identifies

- the types and numbers of morphosyntactic categories (e.g., tense, case, number),
- the number and shape of formatives used to encode them, and
- the combinatorics and classifications (conjugations or declensions) of those formatives as utilized in the language.

The I-complexity of a language, in contrast, measures the (inter)predictability among wordforms; it reflects the ways that the enumerative ingredients cataloged by E-complexity are organized into systems of relatedness among classes of words. This kind of complexity has been a central concern of word-based implicative and network approaches (see, e.g., Ackerman et al. 2009, Blevins 2016, Bonami and Beniamine 2016, Bonami and Henri 2010, Bonami and Strnadová 2019, Bybee 1985, Janda and Tyers 2021, LeFevre et al. 2021, Sims 2015, Sims and Parker 2016, Wilmoth and Mansfield 2021, Wurzel 1987). A hypothesis associated with this division is that a language can, in principle, vary limitlessly in terms of its E-complexity as long as these ingredients are organized in ways that lead to low conditional entropy (LCE; a measure of I-complexity) for the networks of relations between words constitutive of the morphological system.[1] One aspect of this approach is that, while words exhibit internal structure, the nature of that structure is not necessarily morphemic, as typically assumed in familiar generative frameworks. Rather, word structure is defined by discriminability between (classes of) words and the patterns produced by distinctive arrangements of word elements (i.e., segments, suprasegments, and periphrastic constructions) that cohere into systems that constitute language-particular systems.[2]

The central explanatory value of systemic organization for morphological phenomena and learnability makes modern linguistic analysis a beneficiary of the early insights of paradigm-oriented thinkers like Paul, Kruszewski, and de Saussure.[3] For example, Kruszewski (1995) viewed the morphological system as facilitating two fundamental aspects of language usage: *reproduction* was the more or less faithful utterance of stored lexical representations, namely, fully derived and inflected wordforms and their penumbra of variants, while *production* was the utterance of novel wordforms licensed by the analogical inferences intrinsic to networks of related words.

---

[1] Of course, LCE is likely only one, if important and newly explored, dimension guiding morphological organization.

[2] It is important to note that standard morpheme constructs are subsumed under the discriminability view, since the presence of a morpheme obviously counts as one strategy for distinguishing one (class of) word from another. See Ramscar et al. (2018) for discussion of discriminative learning in general and Caballero and Kapatsinski (this volume) for specific discussion of morphemic cues in discriminative learning.

[3] See Blevins (2016) for a detailed review of this tradition and its modern development under the label of Word and Pattern Morphology.

> [E]very word is connected by twofold bonds: by innumerable ties of similarity with its relatives according to sounds, structure, or meaning and by equally numerous ties of contiguity with its various fellow travellers in every possible kind of phrase. A word is always a member of certain nests or systems of words and at the same time is a member of certain series of words. This explains the ease with which we memorize and recall words. Moreover, these properties of words make it possible for us not to have to resort to straight memorization every time. It is sufficient for us to know words like *idu* ['(I) am walking'], *idës* ['(you sg.) are walking'], or *vedu* ['(I) am leading'] in order to *produce* the new word *vedët* ['(he) is leading'], although we may never have heard it before. In the majority of cases we can not say with certainty which words we have learned from other people and which we have produced ourselves; in the majority of cases, as in the above cited examples, parallel forms make it possible to produce only one form, regardless of who is producing it. For this reason W. von Humboldt early on pointed to the perpetual creativity of language. (Kruszewski 1995: 97)

Kruszewski here suggests that the production of a novel inflected form for the Russian verb VESTI 'to lead' is guided by knowledge of other forms of VESTI as well as other inflected forms of the different verb IDTI 'to go.' This represents, according to him, a clear example of an essential challenge presented to theory for language analysis, namely, the "perpetual creativity of language."

Familiar structuralist linguistic theories have operated with a misleadingly "combinatoric" conceptualization of parts and wholes: wholes are of theoretical interest to the degree that they permit the identification of parts that can be recombined algebraically to recompose them with little or no remainder. The whole as representing a distinct level of analysis is foreign to this conception but is central to efforts to understand systemic organization: the internal structures of wholes serve to discriminate wholes from one another, and the networks of relatedness patterns defined by these wholes constitute the analyzable organization of the system.

Significantly, this latter tradition, which developed in parallel with the more familiar post-Bloomfieldian structuralist, morphemic approach,[4] displays conceptual and analytic affinities with research in the "developmental sciences,"[5] where the fundamental constructs guiding explanation include "complex adaptive systems," "systemic organization," and, more generally, a focus on describing and understanding the dynamic interplay between parts and wholes on different interdependent levels that both constitute and define the organization of systems in both nature and culture.[6]

---

[4] See Embick (2015) for a detailed discussion and defense of this "piece-based" conception of morphology.

[5] See Moore (2006) for an overview.

[6] See Ackerman and Nikolaeva (2014), Corning (2018), Hood et al. (2010), Jablonka and Lamb (2014), Laland (2018), Oyama et al. (2001).

Table 8.1 *Comparison of esoteric and exoteric situations.*

Property	Esoteric situation	Exoteric situation
Total community size	smaller	larger
Adult language contact	lower	higher
Learner population	primarily children	contains significant number of adults
Social stability	higher	lower
Communally shared information and traditions	higher	lower
Morphological correlate	higher E-complexity	lower E-complexity

Segueing to the second question (1b) concerning the types of explanation invoked to account for E-complexity differences across languages, work in (typological) sociolinguistics has hypothesized that such differences may correlate with aspects of social structure: languages spoken by large, diverse populations are claimed to be morphologically simpler than those spoken by small, close-knit ones (Kusters 2003; Perkins 1992; Thurston 1987, 1992; Trudgill 2009, 2011, 2016; Wray and Grace 2007). Adopting the terminology of Wray and Grace (2007), we refer to the former as *exoteric situations* and the latter as *esoteric situations*. See Table 8.1 for a summary of the characteristic properties of each.

We will contrast two basic categories of explanations about the relationship between the esoteric and exoteric state of a speech community and the E-complexity of its morphology. The first category of explanation can be referred to as *adaptationist*.

(2) Correlations between social and linguistic types are a matter of adaptation: some language types are "fitter," and therefore selected for, in certain social environments.

Amundson (1996: 25), in developing a more catholic conception of explanation in evolution, identifies the adaptationist strategy as a primary informing hypothesis with a long history:

> To be sure, adaptationists admitted that organs and body parts exist which have no known adaptive purpose. The universal stance on these items might be called the principle of *presumptive adaptation*: Never infer a lack of adaptation from the lack of knowledge of adaptation, because it is always more probable that an unknown adaptive purpose exists than that no purpose exists. The presumption should be that the trait is adaptive, and that eventually its purpose would be discovered.

The primary exemplar of this category of explanation which we will consider here is the *Linguistic Niche Hypothesis* (LNH) (Dale and Lupyan 2011; Lupyan and Dale 2010, 2015, 2016):

> [L]anguages adapt to the learning constraints and biases of their learners. (Dale and Lupyan 2011: 1)

That is, the adaptationist explanation for the observed relationship between social structure and E-complexity is that **both** morphological simplification **and** complexification reflect adaptation to the different learning capacities of L2 and L1 learners in different social situations.

As Lupyan and Dale (2010: 1) put it:

> Our findings indicate that just as biological organisms are shaped by ecological niches, language structures appear to adapt to the environment (niche) in which they are being learned and used. As adults learn a language, features that are difficult for them to acquire are less likely to be passed on to subsequent learners. Languages used for communication in large groups that include adult learners appear to have been subjected to such selection. Conversely, the morphological complexity common to languages used in small groups increases redundancy, which may facilitate language learning by infants.

It is important to emphasize that an adaptive explanation is compatible with four hypotheses: it could explain both simplicity and complexity, only simplicity, or only complexity. It could also, of course, extend to none of these alternatives.

This perspective more broadly embraces a popular and previously prevailing analytic stance concerning the role of external forces on the modification of existing structures. Amundson (2005: 127) refers to this as *the adaptive rule of reconstruction* and formulates it as follows:

> The adaptive rule of reconstruction: Identify ancestral characters and selective forces such that the forces might have caused populations that possessed the characters to diverge into the descendent [*sic*] forms.

In effect, this strategy, characteristic of the Modern Synthesis in biological evolution, has been adopted in other fields that attempt to explain observable change in evolutionary systems: the operative notion is that factors external to the object of change both motivate and shape that change.

The second class of explanations we consider concerning the relation between complexity and social conditions is *neutral*:

(3) Independent of any forces of selection, random variation (evolutionary "drift") can cause E-complexity to increase.

Existing examples of such explanations for sociolinguistic typological patterns can be found in Ehala (1996), Kauhanen (2017), Lass (1997), and Trudgill (2016), inter alia. In the more general context of evolutionary systems, one formulation of this kind of explanation is offered by McShea and Brandon (2010: 4):

> In any evolutionary system in which there is variation and heredity, there is a tendency for diversity and complexity to increase, one that is always present but may be opposed or augmented by natural selection, other forces, or constraints acting on diversity or complexity.

That is, with respect to the observed correlation between social structure and E-complexity, increasing complexity may be the default state of evolutionary systems. This means that no additional explanation is necessary to account for increasing E-complexity in a given language, beyond whatever other contingencies may obtain. In this connection it is important to observe that while increasing E-complexity may have default status, the particular organization of the resulting system (i.e., its I-complexity) may be guided by both internal properties of particular systems as they co-evolve in conjunction with learnability considerations. In other words, the factors responsible for elaboration or simplification in E-complexity may be quite different from the factors responsible for the emergent organization associated with I-complexity. This point is compellingly illustrated in Parker and Sims (2020), where it is shown that the mere enumeration of elements constitutive of, for example, Russian's inflectional morphological system does not provide insight into the important dimension concerning how these elements cohere, let alone why they might cohere in the ways that they do. They conclude (p. 50):

> This suggests that the system as a whole is not simply a function of the complexity of its parts. It is instead a product of the way the parts are distributed – that is, how the component elements are related. This should hardly be a surprise, but the data in this paper highlight that these sorts of local relations, and how they lead to complexity in an inflection class system (or don't!), are at least as important to focus on as the complexity of the system overall. To the extent that languages universally or predominantly exhibit low systemic complexity, the question becomes why. At a broad level, the answer likely has to do with learnability (Ackerman et al. 2009), but to get beyond general formulations of this idea, it will be necessary to dive into the learnability of specific inflection class configurations, and to carefully examine local relations among the component parts of individual inflection class systems.

This can be interpreted as suggesting the importance of distinguishing between E- and I-complexity: E-complexity as derivable from the World Atlas of Language Structures (WALS)[7] provides inventories of morphosyntactic distinctions and their formal exponence, but these alone are simply the ingredients that get organized into the language-particular systems that distinguish a language's morphological organization (i.e., I-complexity). Thus, any hypotheses concerning the relative influences of neutral or adaptationist factors need to clearly identify the scope of influence with respect to E- and I-complexity. For example, it may be that neutral factors influence the E-complexity of a language, while the organization of the resulting elements arises from some adaptationist considerations such as learnability, as mentioned in the preceding quotation.[8]

---

[7] See Section 3.1 for discussion about the limits of what kinds of questions WALS can usefully address.

[8] It is also worthwhile in this connection to consider the valuable reflections contained in Chapter 10 of Bentz (2018). Of particular interest is the recognition that esoteric situations are often characterized by multilingualism, so that contact conditions and the influence of second language learning associated with exoteric situations is not necessarily associated with morphological simplification, as discussed in Meakins et al. (2019).

In this connection, it is important to observe that neutral explanation and, more generally, non-adaptationist perspectives can be seen as complementing rather than replacing adaptationist speculations about specific developments and can themselves be seen as guided or biased by the internal dynamics of the specific systems (Arthur 2004, Riedl 1977, Whyte 1965). Amundson (2005: 127) argues that this system-internal perspective on change and possibility for novelty is the source of fertile reappraisals of adaptation as the single factor of change. He refers to the basic strategy as *the generative rule of reconstruction* and formulates it as follows:

> The generative rule of reconstruction: Identify an ancestral ontogeny that can be modified into the ontogenies of the descendent groups.

What is crucially distinctive here is the focus on ontogenies of development: the mechanics of how a system is organized and operates to yield effects over time.

In sum, the alternatives of adaptive and neutral explanation (with the latter supplemented by considerations of internally guided possible trajectories of change) provide the broader context of competing explanatory resources: while the former is often functionalist in nature, the latter is structuralist, following the traditional distinctions delineated in Amundson (2005).

Our aim in this chapter is to convey to the reader the nature of a neutral explanation of an evolutionary system's state and trajectory and to convince the reader that this type of explanation is a strictly simpler and more likely explanation of higher E-complexity in esoteric situations than the LNH. To accomplish this, we review in the next section the three defining properties of Darwinian evolutionary systems, why language change qualifies as one, what neutral versus adaptationist explanations for the behavior of an evolutionary system are, and why the LNH is adaptationist. In the third section, we begin by discussing the methodological challenges facing evolutionary explanations in biology, language change, and specifically the relationship between high E-complexity language variants and esoteric communities: a lack of data and a wealth of logically possible explanations with unclear or plausibly overlapping predictions. We argue that addressing these problems requires clearly (preferably mathematically) specified models of hypothesized causal mechanisms (e.g., learning), as well as serious consideration of neutral hypotheses and evidence for them, and that simpler explanations (which will often be neutral) should be accepted over more complex ones by default. In the rest of the section, we offer two such simpler (neutral or more neutral) explanations for the same phenomena as the LNH. First, we point out that the main independent variables of the LNH – population size, structure, and other demographic parameters – have been known for more than a century to critically affect the relative likelihood of neutral versus adaptive explanations of the state or trajectory of an evolutionary system; in particular, at least one neutral force – drift – is substantially stronger in small populations than in large ones, can easily be strong enough to overwhelm selectional factors identified in adaptationist approaches, and should be expected to lead to small populations exhibiting and maintaining traits that, if present in an otherwise identical larger population, would be expected to disappear. Second, we review some recent literature modeling language change as an evolutionary process that investigates (among other things) the effects of social structure on the propagation of harder-to-learn versus easier-to-learn linguistic variants. Together, they suggest that even when there is selection against a linguistic variant (i.e., uniformly for all learners in both more esoteric and more exoteric populations), the structure of esoteric versus exoteric populations could lead to a relative homogeneity of input to learners in esoteric situations – enough homogeneity that linguistic variants that need more observations to be successfully learned are plausibly more likely to arise and persist in esoteric populations than exoteric ones. These two results mean that in the absence of strong (forthcoming) evidence for an adaptive explanation of higher E-complexity in esoteric situations (e.g., a benefit to L1 learning), neutral factors are both simpler and specifically more likely than adaptive ones to explain observations about the evolutionary trajectories of historically small, esoteric populations.

2 Background

In this section, we review the defining properties of Darwinian evolutionary systems and why natural language qualifies as one, offer a slightly more technical exposition of the difference between neutral versus adaptive explanations (with examples from both biology and natural language), and then position the LNH with respect to these alternatives.

2.1 Darwinian Evolutionary Systems

A Darwinian evolutionary system can be defined in terms of three abstract elements (adapted from Lewontin 1970, 1978):

(4) a. A population of replicators: A population of objects capable of replicating themselves more faithfully than not from one time step to the next.
b. Variation: Objects in the population can have potentially distinct traits along one or more dimensions.
c. Selection: Some variants in a population are better at replicating than others by virtue of differences in traits.

A trait value that causes those replicating objects that have it to display higher expected success at replication than those with some other variant of the same trait is "adaptive." Insofar as a trait is adaptive with respect to a particular kind of external environment that an object exists in or there is some internal aspect of the object's replication process that shapes its expected success at replication, that trait is said to make the object "fitted" or "adapted" to its environment or "life cycle."

Mechanistically, a Darwinian evolutionary system can be defined by a population state at some moment in time – a frequency or probability distribution over a set of variant types – and an algorithm by which the population at the next time step is generated from the current one (i.e., a set of mechanisms or processes, in parallel or in some sequence, by which replication occurs). Replication involves two basic types of probabilistic choices: choosing for each object whether it survives and replicates, and for each of those that do, choosing how many copies result and how accurately those copies reflect the originals. A replication process that affects which objects survive or replicate only contributes to creating variation when the probability that an object is chosen for survival and replication as a result of that process doesn't depend on its variant type. Similarly, a causal mechanism that affects the number or accuracy of copies of an object chosen to replicate is a mechanism of variation if it doesn't depend on the variant type of the object. In contrast, a causal process affecting a population's dynamics is a selection mechanism when its effect on an object's probability of survival, probability of replication, expected number of copies, or the accuracy of those copies depends on the variant type (traits) of the object.

In the context of biology, examples of different kinds of populations of replicators include:

(5) a. Populations of alleles: different values or forms of a gene
  b. Populations of genotypes: different partial or complete genomes
  c. Populations of phenotypes: different combinations of physical and behavioral traits of an organism

The question of which of these is the most appropriate "unit of selection" can depend on theoretical commitments about biology or evolutionary theory, what scientific question is being addressed, what method has been chosen, or what data are available. Examples of variation mechanisms include the following:

(6) a. Random choice of which organisms die and which reproduce independent of each organism's variant type (drift, discussed in the next subsection)
  b. Random mutation of alleles during replication
  c. Random migration to or from other populations

Some examples of ways that variant types can differ in terms of fitness (i.e., selection mechanisms) include the following:

(7) a. Probability of survival (viability selection)
  b. Probability of reproducing (e.g., sexual selection – the probability of finding a mate)
  c. Expected number of offspring per reproductive event (fertility selection)

An important type of selection that crosscuts classification by biological life stage – and is particularly relevant to cultural evolution – is *frequency-dependent selection*, where the fitness of an individual with a given trait is a function of the relative frequency distribution over traits in the population; we discuss this more below in the context of language. Finally, note that in biology every variation and every selection mechanism listed here is capable of causing the frequency distribution over traits to change, and every variation and every selection mechanism can cause a trait to disappear from a population, but only some variation mechanisms (e.g., mutation or migration) can introduce a previously absent trait, and only some variation and selection mechanisms can act to maintain variation within a population.

While the three abstract elements of (4) suffice to define a Darwinian evolutionary system, in both biology and language, populations of replicators typically have structure that affects what the replication process is, how variation is introduced, and how selection filters or amplifies variation in ways that are substantive, scientifically interesting, and particularly relevant to discussion of the relationship between demographic factors and the relative effects of drift versus selection. That is, a population is supposed to represent a set of spatiotemporally bounded and co-occurring individuals that live, compete, cooperate, and reproduce together in the same context. Suppose a population of individuals (modeled or empirically observed) is meaningfully divisible into two or more subpopulations with a limited and potentially nonuniform rate of migration between them. For example, subpopulations of an organism may be subdivided over different social groups (herds, flocks, etc.) and/or multiple locations like isolated meadows or lakes, an island and a mainland, or the islands of an archipelago. If we want to model the dynamics of this population, we can incorporate our beliefs about this subdivision and organization of the population as accurately as we can, or we can idealize over these differences and treat the population as though it were less structured; our motivation may be practical – a lack of data or the desire for a more analyzable model – or theoretical, for example, exploring how much or little population structure affects the dynamics of the whole population and each of its subpopulations. We roughly summarize the effects of population structure below:

(8) a. All else being equal, the lower the rate of migration, the less the dynamics of each subpopulation are affected by others.
 b. The more asymmetric and heterogeneous population sizes, forces of selection, and migration rates are between populations, the more inaccurate it will be to lump the subpopulations together and treat them as a single unstructured population in a single environment.
 c. The higher the average rate of migration, the more symmetric migration is between subpopulations, and the more similar population sizes and forces of selection are across subpopulations, then the more accurate of an approximation it will be to treat this “metapopulation” as a single unstructured population.

We elaborate in the next section on the relationship between population structure, population size, the effects of forces of variation versus selection, and implications for the LNH.

Analyzing language as an evolutionary system involves making several choices. We schematize these choices as follows:

(9) a. Choosing the set of linguistic representations that variants will be drawn from. For example, variants could be different pronunciations of a phoneme, different strategies for expressing a morphosyntactic property, different synonyms for a meaning, different grammatical strategies for encoding a meaning, generally all or part of a grammar concerned with defining "different ways of saying the same thing" (Croft 2000: 31), or distributions over any of these choices of a set of variants.

b. Choosing a replication timescale – individual dyadic communication episodes versus language development. At its most granular, replication can be taken to be the production of a unit of form – possibly with some meaning and in some episode-specific context – followed by the recognition or comprehension of that form by a listener and some update of the speaker's and the listener's representations of what the language is. Alternatively, replication can correspond to an abstract (child) language development event where some speaker-teachers of the existing community are chosen to provide the input to a learner, who then chooses a linguistic variant (or distribution over variants) at the end of the process and becomes a new speaker-teacher member of the population at the next time step.

c. Choosing a relationship between the population of linguistic variants and the population of speakers in a speech community. The basic replicating object can be taken to be a token of a linguistic variant, and each speaker in a speech community at time *t* can be associated with a population of such tokens – interpretable as a distribution of remembered observations (e.g., exemplars) and/or a production distribution over variant types, and a speech community then corresponds to a population of subpopulations (a metapopulation). Alternatively, the basic replicating object can be identified with a speaker and their linguistic representation – for example, a single linguistic variant, a grammar, or a distribution over variants – and a speech community at a particular point in time can be treated as a population.

While the first choice is relatively straightforward, the last two are more complex and interrelated. For the purpose of understanding language change as an evolutionary process, we discuss different combinations of options for these last two choices below and sketch what population structure and variation versus selection mechanisms look like under each such choice. We begin with the most granular choice of timescale and population.

The most fine-grained choices of replication timescale and population take each speaker in a speech community to represent a population of linguistic variant tokens, the speech community to represent a metapopulation, and individual dyadic communication episodes to be the main process by which the distribution of variants changes over time. Each speaker's population of tokens is most plausibly interpretable as a set of variant tokens or distribution over variant types representing what that speaker has observed themselves and others produce to date[9] or some function of such a distribution (Blythe and Croft 2012, Reali et al. 2014, Wedel and Fatkullin 2017, Winter and Wedel 2016). Replication principally involves repeatedly choosing a speaker and listener pair who will interact, choosing what the speaker says, and an update process describing how one or both participants adjust their internal distributions over linguistic variants as a result. Below is a sequence of events describing how this interaction and update process could be modeled:

(10) a. First, select a speaker s with probability $p_u(S = s)$ and a listener l with probability $p_u(L = l \mid S = s)$ from the population of language users.
b. Suppose there are $X = \{x_1 \ldots x_k\}$ different types of linguistic variants, and that the speaker has to date observed $O_s = \{o_1, o_2 \ldots o_i \ldots o_n\}$ tokens, with the variant type of o_i given by $v(o_i)$.
c. Based on the speaker's observations O_s and a learning or inference algorithm, L, the speaker currently has a production distribution $p_s(X \mid L(O_s))$. They choose a single form x^* to produce by sampling from p_s. A simple example production distribution – exhibiting no selection – might have them randomly choose one of their past observations: $p_s(X = x^*) = n^{-1} \cdot |\{o_i \in O_s \mid v(o_i) = x^*\}|$.
d. The speaker produces a token x^* and adds it to their set of observations.
e. The listener perceives the actually produced form as y, where $p_n(Y = y \mid X = x^*)$ describes how noise can cause the listener to perceive y as something different from x^*.
f. The listener arrives at some beliefs $p_l(X = \hat{x} \mid Y = y)$ about what the speaker actually produced. For example, the listener might reason Bayesianly by combining y with a prior model of what the speaker is likely to have intended to produce p_s' and a model of the noise distribution p_n as $p_l(X = \hat{x} \mid Y = y) \propto p_n(Y = y \mid X = \hat{x}) p_s'(X = \hat{x})$.
g. Using this distribution $p_l(X \mid Y = y)$, the listener chooses some estimate $\hat{x}$ according to a decision rule (e.g., choosing the $\hat{x}$ that maximizes $p_l(X = \hat{x} \mid y)$) and adds it to their own set of observations.

In sum, an interaction between a speaker and a listener leads to production of a form by the speaker which in turn causes a token of some linguistic

---

[9] Note that these could be taken to be perfect or lossy representations of such observations; if they are lossy representations, then the lossy compression and/or noise process by which observations are modified is part of the replication process.

variant x^* to replicate in a "population" of observations associated with the speaker, and after potential modification by noise, perceptual/comprehension processes, and a learning process, to replicate in a "population" of observations associated with the listener.

In order to explicitly represent heterogeneity in types of learners – for example, children versus native adults, contact between language varieties, L2 learners, each with some different initial distribution over observations or learning process – we can specify a rate at which a speaker-listener is added or removed to the population of language users, and a distribution over what kind of speaker-listener is added or removed.

Forces of variation and selection here are determined by

(11) a. the probability distribution over which pairs of individuals are chosen to be speaker-listener pairs,
  b. the probability distribution over what a speaker intends and actually produces,
  c. how production of a token affects the speaker's population of variant tokens,
  d. the probability distribution over what a listener perceives and/or comprehends given what the speaker produced,
  e. how a listener's beliefs about what the speaker said and/or meant affects the listener's adjustment of their population of variant tokens, and
  f. any other details about memory and inference processes specifying how observed tokens of linguistic variants are stored and shape future inference and decision-making of a speaker-listener.

That is, if the probability that any pair of individuals are chosen to be speaker and listener does not depend on the variants of the pair (or distributions over variants of the pair), then that aspect of the replication process would contribute to variation but would not involve selection; similarly, if what the speaker produces, how accurately it is produced or perceived, or how it affects a listener's future inference or production behavior does not depend on the variant of the speaker or listener, then those aspects of replication contribute to variation but would not involve differential selection of some linguistic variants over others.

Other examples of ways in which variants could be differentially selected include the following:

(12) a. Some variants may be more likely to be misheard (Ohala 1993) or misunderstood by listeners or be more likely to vary or be misproduced by speakers.
  b. If speakers and listeners have distributions over linguistic variants, then a speaker may preferentially produce some variants over others if they vary in terms of their estimated sociolinguistic utility (signalling, for example, group identity or prestige) or in terms of their estimated communicative utility (in the sense of, e.g., Lindblom 1990). This production preference over variants could depend on the speaker's own distribution, the speaker's model

of the listener, other communicative and social aspects of the situation, or generalizations the speaker may have made from past experiences, including, for instance, their estimate of the probability distribution over variants of other individuals in the speech community. Note that most of these possibilities are examples of frequency-dependent selection.

c. Listeners may differentially weight or discount a speaker's produced variant in updating their own linguistic variant or distribution over variants in a way that depends on the speaker's produced variant or the listener's variant. This could be caused by, for example, the sociolinguistic properties of the variant, the listener's distribution over variants, or the listener's estimate of the distribution among other individuals in the speech community. Again, some of these possibilities are examples of frequency-dependent selection.

Note that as long as the conditions in (4) are satisfied, we have a Darwinian evolutionary system: no one choice of replicator or timescale of replication here is necessarily exclusive of another. In fact, the model setup and mechanisms described above can be interpreted at a coarser level of analysis, where the timescale of replication is still dyadic communication episodes, but the population of interest (in the sense of [4]) is taken to consist of entities (speaker-listener distributions over variants) that *happen* to also be interpretable as populations. (Hence the term "metapopulation.") Here the space of variant types consists of the space of possible speaker-listener states (the space of distributions over linguistic variants), and the replication process describes how each speaker-listener's population state changes after a communication episode, exactly as before.[10] As before, if the probability that two members of the speech community are chosen to interact as speaker and listener depends on their variant types, then that would constitute a selection mechanism. Similarly, if some variant types (population states) are more likely to accurately replicate than others, then that would also be an example of a selection mechanism.

Coarsening the replication timescale, the replication process can instead abstractly describe (child) language development. This involves choosing a set of one or more speaker-teachers from the set of current speakers, for example, by sampling data from each teacher and applying a model of learning, and based on that choice, generating a new speaker with a new linguistic variant or, more generally, distribution over variants. As before, if each speaker is associated with a probability or frequency distribution over linguistic variants, each speaker can be interpreted as a population of linguistic variants and each speech community as a metapopulation, or (equivalently) a speech community can be interpreted as a population whose members are distributions

---

[10] Learning and inference correspond to replication in the sense that, for example, a speaker-listener's updated variant distribution at time $t+1$ after an interaction at time t is a function of the distribution at time t.

over linguistic variants. The essential differences from the previous choice of timescale are that adults are modeled as static, children learn only from interactions with the previous generation of adults, and details necessary for specifying the process and outcomes of dyadic communication episodes can be abstracted over.

This schematization of language change as an evolutionary system is reflected in the iterated language learning paradigm (Kirby 2001).[11] This is a relatively simple model of cultural evolution intended to facilitate investigation of how the cumulative effect of mechanisms of cultural transmission (i.e., learning) can shape cultural conventions like language over the course of many generations. In the basic version of this model (Griffiths and Kalish 2007), each "generation" consists of one learner. Each agent in a generation learns by observing a sample $O = \{o_1 \ldots o_n\}$ of the cultural behavior (e.g., a set of forms or form–meaning pairs) of the previous generation and then Bayesianly updating their prior beliefs $p(G)$ about what the most likely causes (e.g., underlying grammar(s) or lexicons) of the data they observed are: $p(G \mid O) \propto p(O \mid G)p(G)$. Each agent then samples a hypothesis (grammar and/or lexicon) g from their distribution over causes $p(G \mid O)$ and proceeds to produce data according to their chosen grammar or lexicon for the next learner generation (i.e., according to $p(O \mid g)$).[12] The prior over grammars $p(G)$ reflects the inductive biases of learners; all else being equal, it determines which hypotheses are easier or harder to learn.[13]

The simplicity of this basic form and the use of a Bayesian model of individual inference permits laboratory experiments (for reviews and critical evaluation, see Irvine et al. 2013, Kirby et al. 2014, Mesoudi and Whiten 2008), extensive mathematical analysis (e.g., Griffiths and Kalish 2007) of model behavior and experimental results, and separate manipulation of linguistic representations (e.g., Parker et al. this volume), population structure, and processes of production, comprehension, and learning. Finally, note that Reali and Griffiths (2010) establish a general correspondence between parameter values for a variant of the basic iterated language learning model and the mutation rate of the Wright-Fisher model with drift and K alleles (generalizing beyond the value of $K = 2$ illustrated in the previous section). This result offers a mathematically explicit bridge for connecting the large body of literature on biological evolution to work on iterated learning and forcefully suggests that the arguments about the explanatory burden of neutral versus adaptationist models offered in Section 3 rest on more than just an analogy or abstract similarity between biology and language.

---

[11] See also earlier work by Esper (1925, 1966).

[12] There is typically also a small, fixed, and uniform probability of making a production error.

[13] The more data are available (averaging over possible sets of observations), the less a learner's prior matters and the closer their posterior $p(G \mid O)$ will be to the distribution with all mass concentrated on the teacher's actual chosen grammar. See Griffiths and Kalish (2007: Section 3.1).

2.2 Adaptive versus Neutral Explanations of Variation

Given an evolutionary system, what scientific questions can we ask about it? As summarized by Stephens (2008: 119),

> Population genetics is the study of processes that influence gene and genotype frequencies. It has been obsessed with two related questions: what is the extent of the genetic variation between individuals in nature and what are the factors that are responsible for this variation?

The two questions Stephens identifies apply to any evolutionary system, and answers to them generally emphasize one of (4b) or (4c) more strongly than the other: neutral explanations emphasize the role of mechanisms of variation, while what we have termed adaptationist explanations focus on mechanisms of selection. The question of which type of force is more important (and in what sense) for explaining the extent and dynamics of an evolving population is one of the oldest and most important debates in evolutionary theory.

Below, we exemplify neutral processes in both biology and language: we introduce one of the basic models of population genetics (discussed in more detail in the next section), where a neutral process (drift) is by hypothesis the only force affecting the dynamics of the population, and we discuss an empirical example of complexification in morphosyntax without any obvious or likely explanation in terms of selection.

One of the strongest examples of an answer emphasizing variation mechanisms in biology is "neutral theory" (Kimura 1983), which holds that at the molecular level,[14] mutations and variation we observe are fitness-neutral (or nearly so) and that any given variant's apparent ubiquity within a population (the fixation of a particular variant and the disappearance of alternatives) is more likely a consequence of drift than selection. Drift models the fact that sometimes an organism (or instance of a gene, etc.) in a generation is replicated more or less often than others in the same generation as a result of chance rather than another neutral process – like migration from another population – or a form of selection. That is, drift is one of the simplest ways in which a population of imperfect replicators can imperfectly replicate: a completely random subset of the population is chosen for replication (some potentially more than once), and the rest fail to replicate at all. Figure 8.1 illustrates the hypothetical trajectory of a very small constant-size population ($n = 10$) of gametes of asexually reproducing organisms, where each organism is an instance of one of two possible variants – blank or filled. In linguistic terms, imagine a community speaking a language with two variants in which each speaker uses one of the variants exclusively and the choice of variant is passed directly (but not always accurately) from a single parent to its children (however

---

[14] That is, as opposed to the genetic – a "gene" is an abstraction over molecules.

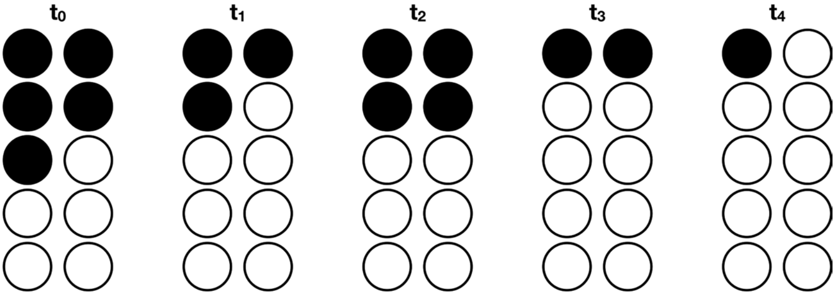

Figure 8.1 A graphical illustration of drift acting on a small population with two variants.

many there may be). The generation at time t_{i+1} is created by sampling with replacement n times from the generation at time t_i: these samples are the members of the new population. Drift is a neutral process because the probability that any particular member of the population at time t_i will replicate doesn't depend on or differ based on the traits of that individual. If one variant was explicitly more likely to be chosen for survival and replication than the other, then the population would be evolving under both drift and selection. To foreshadow discussion in the next section, observe that even though the population in Figure 8.1 started evenly split over both variants, it is quite likely that the population will end up consisting entirely of the blank variant within just a few time steps of t_4 – a complete change in the trait diversity of this population in a handful of generations, all without any forces of selection. In the next section we elaborate on the interplay of drift, population size, and selection, and we discuss implications for adjudicating between neutral versus adaptationist explanations of variation and the LNH.

A linguistic analogue of neutral changes and processes is offered by Trudgill (2011, 2016), who discusses an example from a traditional dialect of southwestern England that underwent morphosyntactic complexification without any obviously adaptive explanation. In this dialect, intransitive infinitives became marked with a wordfinal suffix *-y*, yielding the type of general transitive versus intransitive contrast schematized in (13a). The actual encodings are presented in (13b) and (13c), where the infinitival form required after the modal *can* is either affixless for transitives or affixed by a *-y* for intransitives.

(13) a. *to hit* versus *to runn***y**

b. Can you zew up thease zeams?
Can you sew up these seams?

c. There idden many can shear**y** now.
There aren't many who can shear now.

According to Trudgill (2011, 2016), this typologically unusual marker likely arose as a reanalysis of a phonologically conditioned change. That is, before this innovation arose, all Middle English infinitives had an ending: [-i] for the relevant southwestern dialect variant. We also know that eventually this word-final unstressed vowel was lost in almost all dialects. Appealing to observations of analogous ongoing variation and change in Scandinavian dialects, Trudgill suggests that before this loss was complete, there was a period of variation during which some types of infinitives were more likely to lose this vowel slower – or faster – than others, namely, utterance-finally versus between words (i.e., before obligatory object nouns). While this infinitive marker eventually disappeared everywhere in other dialects, speakers in this southwestern dialect reanalyzed phonologically conditioned variation as an obligatory morphosyntactic marker.

There is no salient reason to think that speakers preferentially produced or learners preferentially inferred – during this transitional period and in this location in England, but in very few other similar contexts – a grammar with explicitly marked intransitive infinitives. Consequently, the propagation and survival of this convention in this speech community is most parsimoniously explainable in terms of neutral processes alone: one or more initial speakers inferred a morphosyntactic reanalysis of phonologically conditioned variation, began producing data consistent with that reanalysis, and other speakers followed suit; eventually it became a convention of that speech community.[15]

2.3 The Linguistic Niche Hypothesis

With a clear sense of the scientific question at hand and two categories of answers, we can now spell out in more detail what makes the claims of Dale and Lupyan (2011) and Lupyan and Dale (2010, 2015, 2016) about the relationship between social situation and morphological complexity adaptationist. The LNH's predictions and the chain of reasoning behind them (Lupyan and Dale 2015) are summarized below:[16]

(14) Predictions:

a. Exoteric condition: The higher the population size and the more area a speech community is spread out over, the less inflectional morphology its language is likely to have.

---

[15] As a reviewer points out, the spread and maintenance of the phonological part of this change might be seen as adaptive if in general apocope is less likely to occur in sentence-final environments. However, our argument here is that there's no reason to think the subsequent development of this phonological alternation into a transitivity marker was an adaptive response to the social, communicative, or learning contexts in which it arose.

[16] It is important to keep in mind that the predictions below reflect the E-complexity properties discussed in Section 1: they ignore how these properties are organized in terms of I-complexity.

b. Esoteric condition: The lower the population size and the smaller the area a speech community is spread out over, the more inflectional morphology its language is likely to have.

(15) Exoteric linking hypotheses:
   a. Increasing population size and the area the population is spread out over is associated with a relatively higher proportion of adult L2 learners.
   b. A higher proportion of adult L2 learners means that there is a smaller portion of the population than there otherwise would be that is likely to be able to successfully learn and use more complex inflectional morphology, likely leading to a trend of decrease in the amount of inflectional morphology, *ceteris paribus*.

(16) Esoteric linking hypotheses:
   a. Decreasing population size and the area the population is spread out over is associated with a relatively lower proportion of adult L2 learners.
   b. A lower proportion of adult L2 learners means that there is no force selecting *against* the propagation of more complex inflectional morphology.
   c. Insofar as inflectional morphology is redundant and a more accessible kind of cue for child language learning than, for example, social or pragmatic reasoning reliant on extralinguistic context, linguistic variants with more inflectional morphology may be learned faster and/or more accurately by children than a language variant with less, leading to a trend of increase in the amount of inflectional morphology in the language, *ceteris paribus*.

As schematized here, linking hypotheses (15b) and (16c) can be understood as describing esoteric and exoteric social situations as different epistemic environments where different types of linguistic variants propagate ("replicate," "transmit," "are learned") more accurately and/or easily by virtue of being more appropriately matched ("adapted") to the strengths and weaknesses of the learner population: exoteric environments select against E-complexity and esoteric environments select for E-complexity. These are what makes the LNH an adaptationist explanation of morphosyntactic variation.

2.4 Interim Summary

Our goal in this section has not been to state the last word or offer definitive technical characterizations of either evolutionary systems generally or language specifically, but rather to illustrate for a linguistic audience the basic structure of an evolutionary process, a basic scientific question one can ask about such processes (viz., the relative burden of neutral versus adaptive explanations), why language change meets the criteria of an evolutionary process, and why the LNH is an adaptive explanation. That is, there are many subtle questions about evolutionary systems that are important to both theory and empirical measurement (e.g., What is the most appropriate unit of selection? When is a trait an "adaptation"? What is the "function" of a trait?) but not to our larger rhetorical goals in the next section: communicating basic results about the strength of selection versus drift in Darwinian evolutionary systems as a function of population parameters like size, the difficulty of clearly identifying selection as the explanation for the distribution of a trait in a population, and why together these make neutral forces a more likely source of explanation for the linguistic typology of historically small speech communities, contra the LNH. Relevant resources for learning about the results and questions in population genetics and evolutionary theory discussed in this chapter include population genetics textbooks (e.g., Hartl and Clark 1997, Rice 2004) and surveys of philosophy of biology (e.g., Hull and Ruse 2008, Rosenberg and McShea 2008, Sarkar and Plutynski 2008).

3 The Burden of Evidence Is on Adaptive Explanations

In this section we discuss two problems facing explanations of variation and change in evolutionary systems: (1) data are generally few and expensive to acquire, and (2) what data we have are often only weakly informative about which of many mechanisms (singly or in combination) caused them. We begin in the first subsection by considering the status of each of these two problems in biology and how it has affected the development and evaluation of theories and explanations there. We then proceed by considering whether similar challenges face the study of language change in general and the relationship between social situation and E-complexity in particular. We conclude that they do and argue for three conclusions about theory development and evaluation for evolutionary explanations of language change:

(17) a. Evolutionary theories of language change need clearly specified models of hypothesized mechanisms affecting replication (e.g., learning).
b. Neutral hypotheses and evidence for them need to be considered and weighed alongside adaptive ones.
c. Simpler explanations should be preferred over more complex ones – especially in the absence of unambiguous data or explicit hypotheses with clear predictions. As discussed below, neutral models are often the simplest explanation.

In the second and third subsections, we argue that there are simpler alternative explanations of a correlation between high E-complexity and esoteric situations that do not require there to be any selective pressures for high E-complexity in general or specifically in esoteric situations, and that therefore the burden of evidence on the LNH is even higher than previously appreciated. Specifically, in Section 3.2 we elaborate on how one of the simplest neutral evolutionary forces – drift – is significantly stronger in small populations than large ones, meaning that we should expect more typological variation across small populations than large ones and that whatever forces of selection are present in them will be blunted or plausibly even overwhelmed by the effects of drift. In the final subsection, we review two recent models of language change that consider (among other things) the effects of high versus low diversity in the language variants of the initial speaker population and of esoteric versus exoteric social network structures. Together, they suggest that the relative homogeneity of input in esoteric social situations relative to exoteric ones means that any linguistic variant (e.g., potentially high E-complexity ones) requiring more observations to learn is more likely to be learned in an esoteric social situation than an exoteric one – crucially without any requirement that learners specific to the esoteric environment favor the more difficult variant or that learners specific to the exoteric environment favor the simpler variant.

3.1 Challenges of Explanation in Evolutionary Systems

Stephens (2008: 119) describes some of the challenges facing attempts to explain variation in an evolutionary system and offers one of the key methods by which the study of biological evolution has made progress:

> Much of the historical, methodological, and philosophical interest in population genetics results from the fact that [its] two central questions – the extent and explanation of genetic variation – have proved extraordinarily difficult to answer. It is impossible to know the complete genetic structure of any species, and there are significant underdetermination problems in figuring out which factors are the relevant causes of evolutionary change, even if one knows a lot about the genetic structure of a population. Despite these difficulties, population genetics has had remarkable successes, and is widely viewed as the theoretical core of evolutionary biology. Significant evolutionary changes often occur over thousands or millions of years. Because of this, it is impossible to observe these changes directly. As a result, understanding the causes of evolution depends crucially on theoretical insights that flow from the mathematical models of population genetics.

That is, in the face of data about genetic variation that were both hard to come by and a variety of hypothesized mechanisms by which that variation could change (rendering most data underinformative), biologists expended great effort in elucidating the space of theories by constructing explicit mathematical models where the presence of different causal mechanisms affecting replication can be toggled on or off, parameters (e.g., population size, mutation rate, strength of selection) can be varied or related to empirical measurements, and the predictions of different modeling assumptions can be compared to each other and what data are available. As elaborated in the next subsection, these formalizations of Darwinian evolutionary dynamics show that drift should be expected to have a strong effect on the evolution of small populations and relatively little effect on large ones.

Mathematically explicit theories of evolution were not enough, however; they needed to be complemented by careful scientific reasoning about available evidence and consideration of available explanations. Historically, one of the main arguments of critics of adaptationist explanations in biology (prominently, Gould and Lewontin 1979) was that researchers offering such explanations for empirical phenomena often failed to seriously investigate or consider the relative evidence for neutral explanations of the same phenomena and accepted the apparent sufficiency of an adaptationist explanation on the basis of weak empirical evidence. Nevertheless, it is commonly noted (see, e.g., Pigliucci and Kaplan 2000) that one of the legacies of Gould and Lewontin (1979) over the last few decades has been an improvement in standards of evidence for adaptive explanations in evolutionary biology.[17]

In sum, in the face of insufficient empirical data and a complex hypothesis space full of theories making overlapping predictions, evolutionary biology proceeded in two directions: (1) by clarifying mathematically the nature of each hypothesized causal mechanism affecting replication, identifying what data it predicts as well as how it compares or combines with other mechanisms, and (2) by holding adaptive explanations of empirically observed variation and change to a higher standard of evidence.

What is the situation facing language? Generally speaking, data about variation and change are at least as hard to come by as in evolutionary biology and at least as indeterminate with respect to ultimate causes. In fact, even our theories of causal mechanisms affecting replication and their relative frequency, strength, and interaction are in their infancy: insofar as we have explicit models of language learning, comprehension, or production in individuals, we have only begun to examine how these function at population- and historical-scales, how they interact, how or when each should be expected to be strong or weak, or how they relate to sociolinguistic factors (e.g., Blythe and Croft 2021, Kocab et al. 2019, Raviv et al. 2019, Roberts and Sneller 2020).

Finally, while we discuss some recent work in the next section that has begun to address these problems, few to our knowledge have yet examined detailed or realistically complex linguistic representations. Turning to evidential standards for adaptive versus neutral explanations in language, Lass (1997) argues that much functionalist work (including in the context of morphology) assumes that there is some teleological force of change in the direction of transparent (one-to-one) form–meaning mappings, motivated by the putative need to resolve the absence of clear function–form organization whenever this occurs. Discussing a representative proponent of this principle, Lass, who dubs it *the mind shuns purposeless*

---

[17] Note also that one of the important roles of mathematical models of the neutral theory of molecular evolution (Kimura 1983) was providing a null model for inferring the presence and strength of selection from molecular data.

variety (mspv), provides several examples where the principle appears to obtain, while demonstrating that there are many others where it does not. What is the status of such an adaptationist principle, given such a state of affairs?

> If we invoke mspv only for good outcomes, and allow bad ones to be not counterexamples but simply non-instantiations of something 'tendential' in the first place … the mspv explanation is invincible, and therefore uninformative …. This suggests that either 'the mind' doesn't behave this way (the obvious conclusion); or that the variety is not purposeless, or is at least neutral, in the sense that preference and dispreference are both arbitrary. (Lass 1997: 344)

The lesson of present relevance for the analysis of language change is one of caution: adaptive explanations require careful elucidation in each instance and should be distrusted without such specification.[18]

Accordingly, while empirical data about language change and its mechanisms continue to accumulate, we argue that adaptive explanations of language change need to clearly describe hypothesized mechanisms and weigh evidence for their hypothesis with evidence for neutral ones. In cases where, on the one hand, data are sparse and relatively indeterminate and, on the other hand, we do not have a clear sense of what the space of hypotheses is or what predictions they make because mechanisms of language change have not or are only beginning to be explicitly formulated and analyzed, the principle of Occam's razor suggests that we should prefer simpler explanations over more complex ones. Insofar as neutral explanations of available data typically require fewer and/or weaker assumptions about what drives evolutionary change than adaptive ones do, they ought to be regarded as a priori more likely.

In the specific case of the LNH's adaptationist hypothesis about E-complexity and esoteric communities, the situation outlined above for language with respect to data, theory, and consideration of alternative neutral explanations is even more pronounced. On top of the uncertainty about the relevance of E-complexity versus I-complexity for understanding morphological complexity expressed in Section 1, it is still unclear whether there is a veridical correlation between E-complexity and esoteric communities.

---

[18] There is another lesson which applies to the general information-theoretic implicative framework which guides this chapter. Morphological organization analyzed in terms of low conditional entropy between words does not mean that such systems strive toward lower and lower conditional entropy values: in fact, conditional entropies can increase over time. Languages simply utilize whatever forms arise and (re)organize them into systems of greater or lower conditional entropy, as long as they retain enough transparency to be learnable. In other words, following Lass's observation, changes are not driven by tendentious (dis)preferences: there are, to our knowledge, innumerable (re)organizations compatible with the need to be learnable. Maiden's documentation (Maiden 2018) of morphological perseverance, that is, the maintenance of complexity where simplification would be expected owing to functional considerations, reinforces Lass's wariness regarding claims about change being motivated by impressionistic learnability considerations.

First, the statistical correlations proposed by the LNH and others are based on language data drawn from the World Atlas of Language Structures, or WALS (Dryer and Haspelmath 2013). WALS was constructed to support typological investigations by linguists. While it has proven its worth in that domain, many of its properties make it less well-suited for use in large-scale regression models. The information in WALS was collected over many years by various research groups for different purposes, and this naturally has led to large variation in quality and detail. Unavoidably, coding errors have crept in. For example, Rubino (2013) lists Nandi as a language that exhibits productive reduplication, but the source cited for this information is actually describing Kinande, an unrelated language. In many other cases, the coding is technically correct but obscures important differences between languages. Take the entry for number of nominal cases (Iggesen 2013), a linguistic property that is clearly an aspect of E-complexity. The number of cases in a language would seem to be straightforwardly quantifiable as an integer. But, in WALS it has been arbitrarily discretized into eight categories. This makes spatial visualizations simpler but complicates the use of this feature in further statistical analyses. More deeply, individuating and enumerating cases is not without problems, even setting aside the issue of quantization. English is listed as having two cases; while this is not wrong, exactly, case marking in English is marginal at best and has a very different status in the grammar than it does in, say, Modern Irish. Also, cases with non-syntactic functions (like the vocative) were left out of the counts, as were genitives that agree in person or number with the possessed noun. These choices are justified and documented in the relevant WALS chapter, but subtleties like this get lost when many different features are combined into a single large statistical model.

Second, most of the demographic information we have is only weakly informative about the LNH and its object of explanation: most measurements that we have are limited to recent history, and languages with historically small speech communities are in general the ones for which we are likely to have the least data, especially historical data. Finally, the problems with each of these sources of data compound each other when correlations between them are examined: only some fraction of demographic data about a speech community is likely to be associable with relevant historical descriptions of the language with enough detail to draw conclusions about E-complexity.

Turning to theorized mechanisms and empirical predictions, the hypothesis embodied by the LNH about the relationship between E-complexity and esoteric communities has little empirical data and no explicit models indicating

(18) a. why high E-complexity could or should be expected to facilitate L1 but not adult L2 learning,
b. why ease of learning among children of higher E-complexity variants is at the expense or exclusion of later acquisition or use of lower E-complexity variants,

c. that this pressure *for* high E-complexity everywhere there are L1 learners could plausibly be, or is in fact, weaker in exoteric situations than a pressure *against* high E-complexity, or
d. how the predicted observations of such forces compare qualitatively or quantitatively (i.e., in relative strength) with neutral explanations of variation and change.

That is, without an explicit account of how high E-complexity ought to facilitate L1 learning, it is difficult at minimum to understand its predictions or to evaluate it against empirical evidence. Second, without one or more neutral models of variation and change, there is nothing to compare either the empirical evidence or LNH predictions against, nor is it clear what the conditions are for hypothesized forces of selection to outweigh the effects of neutral forces – as opposed to being overwhelmed by them, as the next subsection notes is particularly plausible in small populations. Third, to really evaluate or understand the predictions of the LNH, we not only need to see an explicit model of L1 learning and its relation to E-complexity that supports the LNH, but also one of adult L2 learning and its relation to E-complexity. The reason why is that the posited causal mechanism behind the LNH's explanation of esoteric typology isn't actually something unique to esoteric situations – it's something present in *both* esoteric and exoteric situations (child learners) and a relative lack of something present in exoteric situations (adult L2 learners). As a result, any given variation and L1 learning model sufficient to predict selection *for* high E-complexity in all populations where there are child learners could end up predicting that the pressure for high E-complexity should in general prevail relative to any given adult L2 learning model sufficient to predict a preference by them *against* high E-complexity and for low E-complexity. Given the general expectation in evolutionary systems (elaborated in Section 3.2) that drift is in general much stronger in small populations, and therefore only relatively strong forces of selection should be expected to reliably shape their evolution and so be a reasonable explanation for the typology of small populations, this concern is doubly important for the LNH. Altogether this means that to be compatible with the full range of the LNH's predictions about E-complexity and social situation, any model of L1 learning offered in support of it that is sufficient to predict selection for high E-complexity in esoteric situations must also be weak enough relative to the selection pressure of a model of adult L2 learning sufficient to predict selection against high E-complexity in exoteric situations. Accordingly, understanding and evaluating the predictions of any model of L1 learning offered in support of the LNH's predictions about esoteric situations is partially dependent on what model of adult L2 learning is offered in support of the LNH's predictions about exoteric situations.

Turning to standards of evidence, Dale and Lupyan (2011) and Lupyan and Dale (2010, 2015, 2016) spend little time considering or weighing neutral alternative explanations for the relationship between E-complexity and esoteric communities. Lupyan and Dale (2010: 8) do acknowledge drift briefly as an alternative hypothesis, but do not elaborate or discuss the relative strength of the evidence for it. Dale and Lupyan (2011) offer no investigation or discussion of neutral mechanisms at all or what they would predict about either their agent-based simulation or their empirical investigation. Lupyan and Dale (2015) discuss "drift"; however, they do not use the term to describe a neutral random sampling-like process affecting which elements of a population survive and replicate,[19] but instead to describe two separate phenomena in an agent-based simulation of theirs. First, they use it to describe a linguistic analogue of *allopatric speciation*: when a population splinters into two or more geographically isolated populations, the populations may evolve along different evolutionary trajectories – "drift apart," in this sense of Lupyan and Dale (2015)'s usage. This kind of divergence in evolution between geographically isolated populations is *not* synonymous with drift, the evolutionary force. Rather, it can be a consequence of neutral forces like drift, differing selection pressures in different environments (as in the simulation of Lupyan and Dale (2015)), or some combination of both. In the case of Lupyan and Dale (2015)'s simulation, the divergence in evolution of isolated groups is a consequence of a non-neutral migration model that preferentially keeps agents whose language variants are sufficiently similar together, geographically varying selection pressures, and the selective force their second usage of drift refers to. This second sense of "drift" in Lupyan and Dale (2015) refers to an accommodation-like mechanism in their simulation whereby speakers adjust their linguistic representations to more closely match the average value in their local speech community – a frequency-dependent selection mechanism, not a neutral one. Finally, although a sidebar in Lupyan and Dale (2016) correctly indicates that "drift" in the context of evolutionary theory refers to random sampling-like effects on which individuals survive and replicate, the main text only uses "drift" to refer to the process and effects of a linguistic analogue of allopatric speciation. Neither Lupyan and Dale (2015) nor Lupyan and Dale (2016), then, discuss or evaluate neutral explanations.

In the next two subsections, we discuss alternative explanations that do not require the assumptions that there are different kinds of learners or that there is any selective pressure for high E-complexity specific to esoteric situations. These alternatives predict that small, esoteric populations should still be expected to display a greater degree of variation (Section 3.2) than large, exoteric ones. Even if a small population is subject to selection,[20] drift is more likely to be the cause of evolutionary changes (Section 3.2), and small populations are more likely to permit difficult-to-learn variants (if they exist)

---

[19] I.e., what "drift" conventionally means in the context of evolution.
[20] Regardless of whether it is specific to small populations.

to persist or become common than in large, exoteric populations that are otherwise comparable (both Sections 3.2 and 3.3).

3.2 Drift Is a Powerful Force on Small Populations

Recall the basic structure of the LNH's explanation for the relationship between social situation and E-complexity presented in Section 2.3: the independent variables whose value or direction of change precedes all other steps in the causal chain of the LNH are demographic variables like population size. Among the kinds of forces – drift, migration, mutation, and selection – commonly examined in population genetics, drift is known[21] to be much stronger in small populations than large ones, and – for biologically plausible mutation rates and relative fitnesses – to be much more powerful than mutation or selection in small populations and negligible in large ones.

To illustrate this and its consequences for reasoning about what explains the observed state of an evolutionary system, consider again the hypothetical population discussed in Section 2.2 of a small population shown in Figure 8.1. This is a possible evolutionary history of a population of ten individuals with two possible trait types[22] over five generations in a variant of one of the basic models of population genetics: the Wright-Fisher model with drift, but no mutation, no migration, and no selection. Recall that this means that the generation at time t_{i+1} is created by sampling with replacement n times from the generation at time t_i, and this collection of samples constitutes the population at time t_{i+1}: the probability of any individual in the population at time t_i surviving and producing one replicant does not depend on the variant type of that individual and is uniform across the population.

Figure 8.2 illustrates what happens as we increase population size in this variant of the Wright-Fisher model: each graph shows the trajectories over twenty generations of ten different populations that all start out with a 50/50 distribution over the two trait types. The y-axis summarizes everything about the state of a population at a particular point in time in terms of the proportion of that population with one of the two trait values.[23] Figure 8.3 is the same, but for 1,000 generations of evolution. As population size gets larger, it's clear

---

[21] See, for instance, the population genetics textbooks Hartl and Clark (1997) or Rice (2004).

[22] To be more specific: in biological terms, these are individuals with one allele per gene (they are "haploid") who reproduce asexually, and we are modeling the evolution of one locus ("gene") that can take on exactly one of two possible values ("alleles") and whose evolution is, by assumption, independent of all other loci in the organism's genome. As in Section 2.2, this corresponds to a community in which each each speaker uses one of two linguistic variants exclusively and in which the preferred variant is (noisily) inherited by (potentially variable) children from a single "parent" individual.

[23] What was blank versus filled in Figure 8.1 is here variant A versus a.

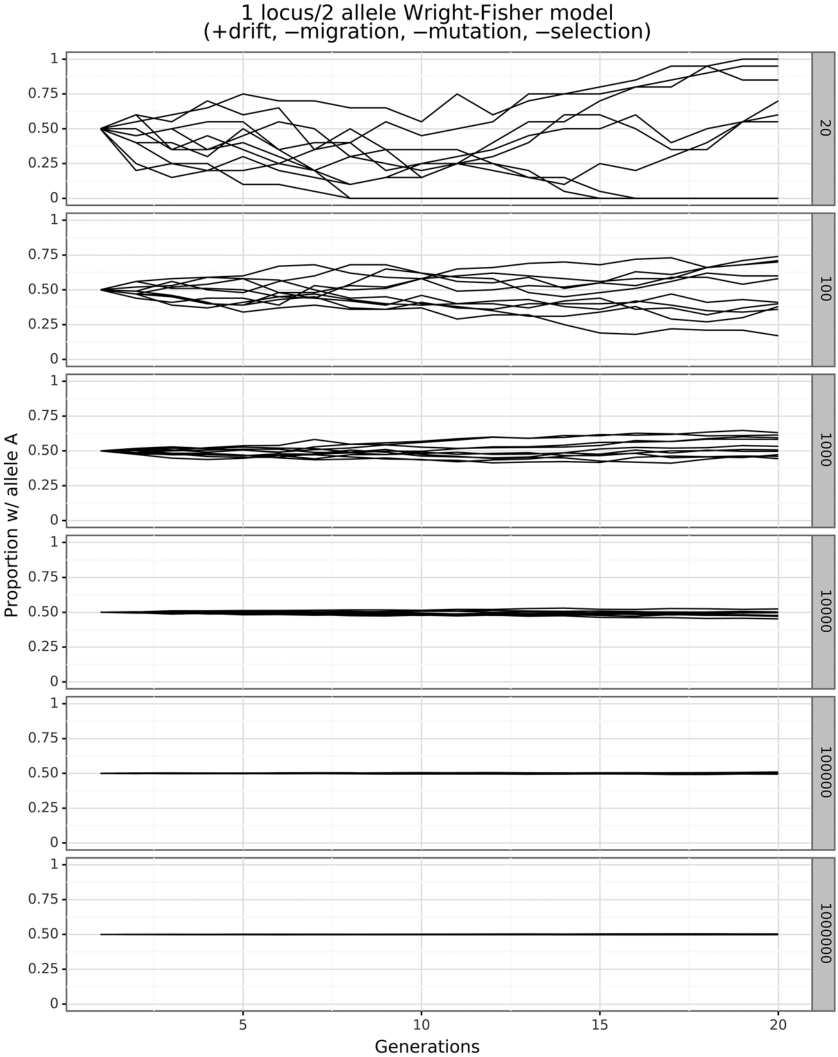

Figure 8.2 Each plot shows the trajectories (under drift alone) over 20 generations of 10 simulated populations with population sizes (indicated on the right) varying from 20 to 1,000,000.

that drift has less and less effect per unit time: drift will take much longer, compared to when the population is small, to push a population's state a given distance from the same starting point.

These graphs illustrate several notable properties of drift as a force acting on a population:

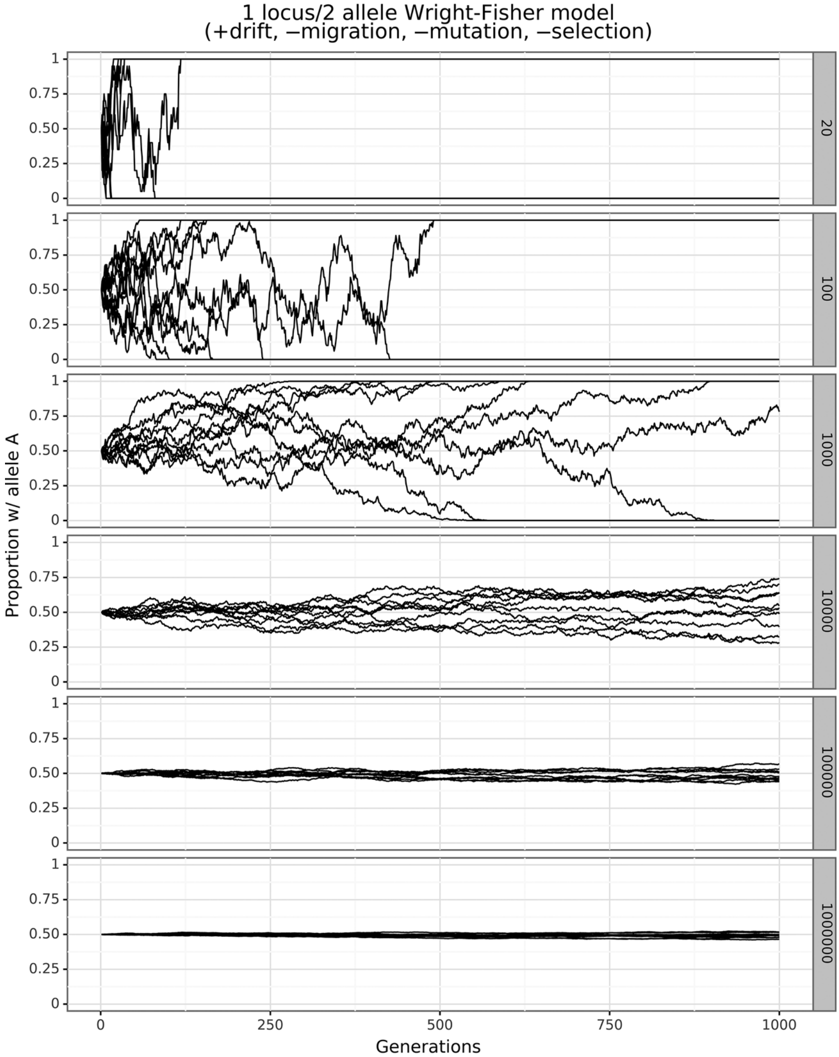

Figure 8.3 Each plot shows the trajectories (under drift alone) over 1,000 generations of 10 simulated populations with the population size (indicated on the right) varying from 20 to 1,000,000.

(19) a. The absolute frequency of each variant undergoes fluctuations that are usually small at each step.
b. With one important exception (19c), fluctuation in one direction is as likely as any other – hence the name “drift.” This is in contrast to other forces, like selection or potentially “directed” neutral forces like asymmetric migration

rates, as may, for example, be the case in a biological context between a small island population and a larger mainland one.

c. Once a population evolving under drift contains no individuals of a variant type, that type will never appear again unless another source of variation (e.g., mutation or migration) reintroduces it.

Within a population, this means that in a small number of generations, drift causing a small number of changes in the absolute frequencies of a small population can cause a large change in relative frequencies: the population in Figure 8.1 started evenly split over both variants, but it is quite likely that the population[24] will end up consisting entirely of the blank variant within a few time steps of t_4. By the same token, the larger a population is, the less effect drift has on the trajectory of the population and the longer it will take for drift to cause one variant versus another to sweep to fixation. Drift also has important between-population effects: small subpopulations of the same species that are relatively separated from each other (due to, e.g., geographic distance or other barriers) will each undergo drift, but do so separately (i.e., in uncorrelated directions). Without the intervention of other forces like high enough rates of migration or similar directed forces of mutation or selection operating in each subpopulation, the members of each subpopulation will likely become more similar to each other than to members of other subpopulations.[25] To summarize: with respect to a single population, the smaller the population, the stronger drift is as a source of long-term change – specifically, loss of variation and increase in within-population homogeneity; with respect to multiple relatively separated and small subpopulations, drift is a force for diversification and divergence between those subpopulations.

In the context of morphology and esoteric populations, the takeaway is that, all else being equal, random fluctuations in replication frequency that are small in absolute number ought to be expected to have a much stronger effect on language change per unit time in a small community than in a large one, and that, all else being equal, drift will cause much more typological variation (including, for instance, some amount of high E-complexity) across small populations on a given timescale than it will across large ones. Note that without any assumptions about selection for *or* against high E-complexity under any circumstances, drift alone should be expected to lead to more variation

---

[24] Because of the simplifying assumption made in this example that each speaker uses only one of the two variants, the population of variants maps directly onto the population of speakers. In general, however, it will be important to keep the distinction between these two populations clear.

[25] If this proceeds far enough for long enough, it can lead to allopatric speciation, referenced in the previous subsection.

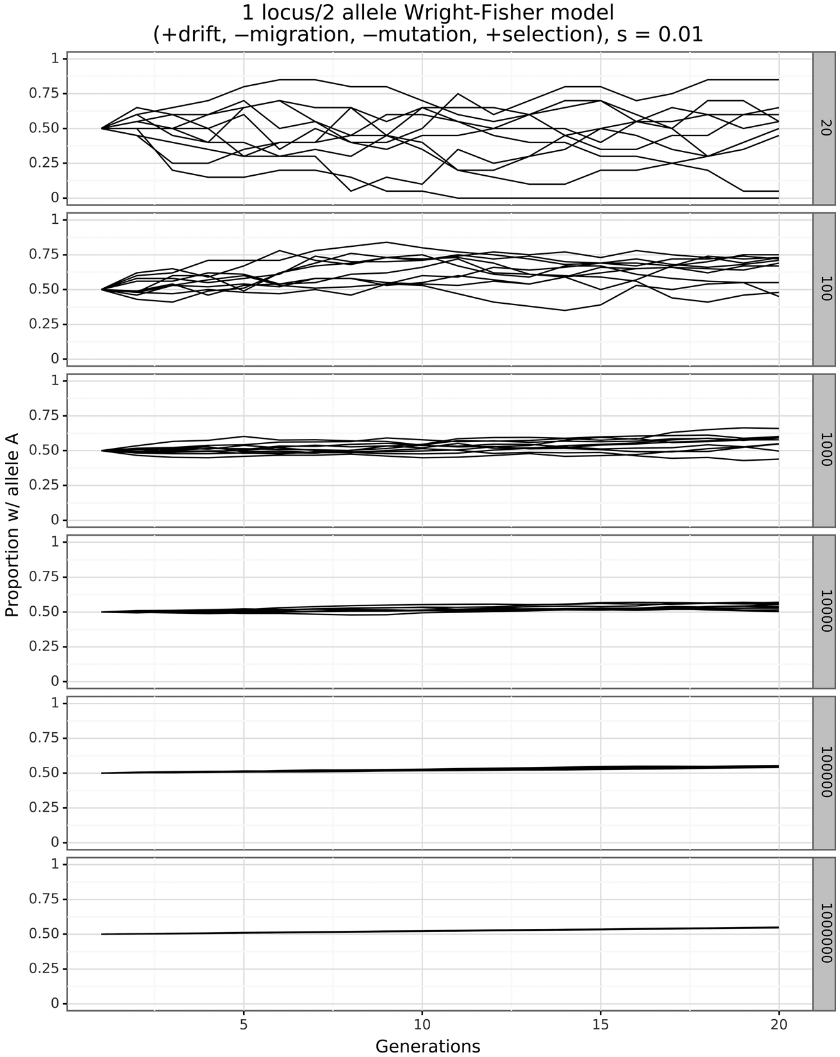

Figure 8.4 Each plot shows the trajectories (under drift and a moderate amount of selection) over 20 generations of 10 simulated populations with the population size (indicated on the right) varying from 20 to 1,000,000.

across a set of small populations at any given moment than across an otherwise comparable set of large ones.

How do the effects of drift and selection interact as a function of population size? Figures 8.4 and 8.5 are similar to the previous pair, except that they now illustrate a Wright-Fisher model with a moderately strong amount of

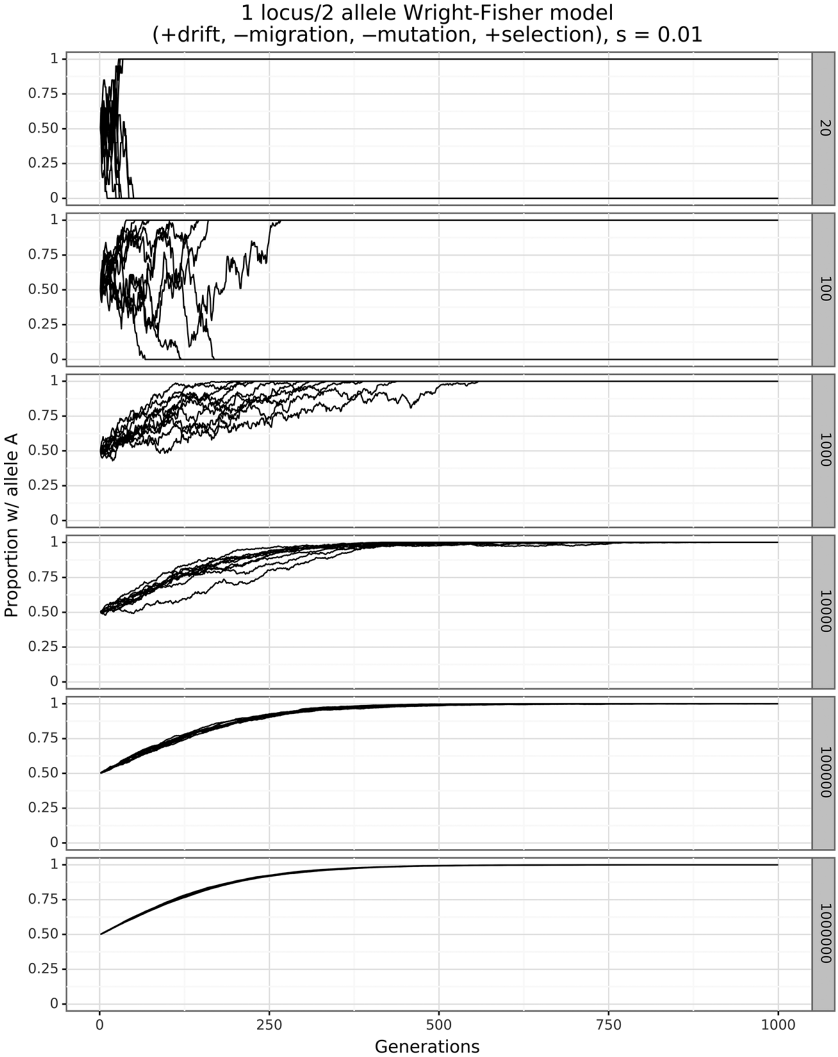

Figure 8.5 Each plot shows the trajectories (under drift and a moderate amount of selection) over 1,000 generations of 10 simulated populations with the population size (indicated on the right) varying from 20 to 1,000,000.

(frequency-independent) selection for one of the two trait values.[26] As population size gets larger, the effect of drift becomes weaker, and the direction and strength of selection becomes clearer.

---

[26] See any population genetics textbook for reference on the relevant calculations.

With respect to the LNH, then, it is plausible that even if high morphological E-complexity *were* clearly and demonstrably advantageous for child learning relative to low morphological E-complexity, the effects of drift in small (esoteric) populations could plausibly mask or even overwhelm it. Generally speaking, it means that whatever forces of selection operate in all populations of speakers, drift should be plausibly expected to cause changes (typological variation) in relatively small populations that selection would be expected to filter out in relatively large ones. Reasoning about how likely this is as a relevant concern for the LNH, or what the general conditions are for this to likely be relevant,[27] requires evidence about the relative strength of drift versus different kinds of selection in language change – such as analysis of an explicitly presented model with mechanisms of drift and both L1 and adult L2 learning, more empirical data on learning, longitudinal data on population size, population structure, and E-complexity.[28]

Finally, these graphs should also drive home the importance of empirical data on how populations change over nontrival stretches of history for reasoning about what explains the typology we see currently. That is, consider the task of trying to determine the strength of evidence for selection in explaining the observed diversity of traits across several populations. Above we have simulated data for several such populations from a model that is an idealized, controlled, and oversimplified representation of biological evolution, and – crucially – we have many longitudinal measurements covering the entirety of a long timespan. In contrast, as noted earlier in Section 3.1, relatively little and sparse diachronic data are available about the linguistic structures, the relative "fitness" of those structures, or the social structures for many of the languages and speech communities in Lupyan and Dale (2010)'s WALS-based analysis. Compare Figures 8.2 and 8.4, but imagine only seeing the state of a few populations from either graph and only seeing one or two points in time for each population. Under these conditions, determining with confidence whether a population is being acted upon by selection or only subject to drift is extremely difficult, and our conclusions should be appropriately qualified and conservative (Smith 2016).

In sum, the main independent variables behind the LNH's adaptive explanation of the relative prevalence of high E-complexity in esoteric situations should also be expected to amplify the effects of a much simpler neutral evolutionary force – drift. Drift is an undirected force that should lead, all else being equal,

---

[27] I.e., for a given model of drift in language change, what counts as a large enough population size for drift to no longer have an appreciable effect on a particular timescale, in the absence of selection or in the presence of a particular kind and degree of selection? Given a particular model of drift, some choice of kind selection, and an empirically plausible degree of selection – whatever that may turn out to be – how large does a population need to be for the effects of selection to likely outweigh the effects of drift on a particular timescale?

[28] We refer the reader to a detailed exploration of several of the external factors of influence on language complexity in Bentz (2018).

to relatively large amounts of variation between small populations in a given span of time, and the relative strength of drift could plausibly overwhelm any effect of selection, if present. This also means that we should expect drift to cause relatively small populations to display linguistic variants that selection would remove in a larger population. Both drift alone and the LNH's two-part selection for high E-complexity in esoteric populations and selection against high E-complexity in exoteric ones are hypotheses that could predict observing more high E-complexity language variants in small populations. However, because drift is a strictly simpler explanation and more likely to explain differences in the evolutionary trajectories of small versus large populations, drift should be regarded as an a priori more likely explanation than the LNH until we have clearer empirical evidence or model-based reasoning to suggest otherwise.

3.3 Relative Homogeneity of Input in Esoteric Populations

In this subsection we review two computational models of language change that manipulate the composition and structure of populations. Both indicate that, if there are difficult-to-learn linguistic variants (where difficulty is uniform across all learners), then small, esoteric populations are more likely to permit these difficult-to-learn variants to persist or become common than large, exoteric populations that are otherwise comparable. They also indicate that this is explainable as a consequence of differences in population size and structure in small, esoteric populations versus large, exoteric ones rather than differences in which forces of selection are operating in esoteric versus exoteric populations.

The first model is a variation on the Bayesian iterated language learning model outlined in Section 2.1. Whereas in the simplest form of this model each learner observes data produced by exactly one teacher who has chosen exactly one grammar as the basis for their productions, Burkett and Griffiths (2010: Section 4) and Dangerfield (2011) consider a more realistic setting where each learner's data come from multiple teachers of the previous generation – and hence from multiple grammars. The task of learning is still reasoning about how likely different causes are to have given rise to the observed data, but now a "cause" is a distribution over grammars rather than a single grammar. Accordingly, where the learners of Griffiths and Kalish (2007) discussed previously have a prior over grammars, learners in this multi-grammar setting have a prior over distributions of grammars. While a technical discussion of the form of this prior is outside the scope of this chapter, all that is important for the present discussion is that this prior has two parameters, a base distribution over grammars G_0, and a concentration parameter α. The base distribution is comparable to the prior over grammars discussed earlier, while the concentration parameter reflects the learner's expectations about both how many distinct grammars are responsible for the observed data and how close their distribution over grammars is to the base distribution:

$0 < \alpha << 1$ indicates an expectation that increasingly many datapoints are produced by very few grammars and where the distributions most likely to dominate are decreasingly close to the base distribution, while $1 << \alpha$ indicates an expectation that increasingly many datapoints are produced by increasingly many grammars, and where the distribution over which grammars those are is increasingly close to the base distribution.[29] As the number of observations increases, the effect of a learner's prior diminishes and their posterior will approach the actual generating distribution; as α decreases, the rate at which this happens will increase. Given the relatively small number of observations per learner in Burkett and Griffiths (2010) and Dangerfield (2011), we are interested in moderate-to-lower values of α. In this parameter regime, the end result of iterated learning is an amplification of biases in the data presented to the initial population of learners.

That is, consider two instantiations of the model from Burkett and Griffiths (2010: Section 4) or Dangerfield (2011: Chapter 5), one where the initial data are consistent with a relatively flat distribution over grammars – an *exoteric* starting condition with a relatively heterogeneous mix of grammars – and another where the initial data are consistent with a relatively peaked distribution over grammars – an *esoteric* starting condition with a relatively homogeneous mix of grammars. Absent some reason to expect an exoteric learner to observe more datapoints overall than an esoteric learner, a learner in the exoteric starting condition receives strictly fewer datapoints per language per unit time than an esoteric learner. This means that for any grammar variant g_{hard} that is more difficult to learn[30] than another grammar variant g_{easy}, exoteric learners will be less likely to end up selecting that grammar (given the same amount of data) than learners in a much more homogeneous population consisting principally of speakers who preferentially use g_{hard}. In the context of the LNH, this means that if high E-complexity language variants are indeed harder to learn for (all or any significant fraction of all) learners, then homogeneity of input in esoteric situations could be sufficient to allow harder-to-learn variants to be more likely to persist than in exoteric situations. Crucially, note that this does not require that high E-complexity be particularly *beneficial* to a type of learner that is specific to the esoteric social situation.

Reali et al. (2014) offer simulation results roughly mirroring the logic outlined above, but with three notable differences from Burkett and Griffiths (2010). First, where the model of Burkett and Griffiths (2010) is comparable to the discrete, nonoverlapping generations Wright-Fisher model of population genetics, where a replication event is synonymous with an abstract child language acquisition event, Reali et al. (2014) uses a model more comparable to the overlapping

---

[29] See Dangerfield (2011) for extensive discussion of the concentration parameter.

[30] I.e., require more observations on average for a learner to assign it a given probability.

generations Moran model of population genetics, where the replication process is comparable to individual episodes of production of a single utterance. Second, while the framework of Burkett and Griffiths (2010) does permit explicit manipulation and analysis of what variants require greater versus fewer expected observations to acquire, Reali et al. (2014) do, in fact, explicitly manipulate the learning difficulty of linguistic variants. Third, where Burkett and Griffiths (2010) model every speaker-teacher from generation t as equally likely to contribute data for each new member of generation $t+1$, Reali et al. (2014) assume a spatialized model where each speaker only interacts with nearby agents. While both Burkett and Griffiths (2010) and Reali et al. (2014) are neutral models insofar as the probability that any given speaker contributes data that influence a listener does not depend on their linguistic variant or distribution over variants, Reali et al. (2014) is both more realistic and specifically permits exploration of the idea that differences in the network structure of who talks to whom in esoteric versus exoteric communities contributes to differences in morphological typology (see, e.g., Trudgill 2009).

In more detail, Reali et al. introduce a kind of spatial structure to communicative interactions and allow the learnability of different linguistic conventions to vary. They simulate a persistent population of communicating agents by placing each agent on a unique node in a type of random graph whose structure allows for gradient exploration of conditions corresponding to esoteric and exoteric social situations: as the number of nodes in the graph (population size) increases, the average number of neighboring nodes increases. Crucially, only agents in nodes that are connected (neighbors) can communicate with each other. As a result, each agent in the esoteric condition tends to have repeated interactions with a small number of speakers who *themselves* tend to have repeated interactions with a small number of speakers (and so on). Accordingly, a linguistic convention requiring relatively more observations to be accurately learned is more likely to perpetuate itself and take hold in an esoteric population than an exoteric one, all else being equal.

The models and results of both papers offer simpler alternative explanations of why small, esoteric populations are more likely to display variants that are harder for some portion of the population to learn than large, exoteric ones. Consider that the LNH's explanation depends on the following:

(20) a. the existence of a force of selection for high E-complexity
  b. the hypothesis that this force is explained by a model of child learning favoring high E-complexity variants and leading to their preferential later use
  c. the hypothesis that this is specifically due to children having an easier time keeping track of redundant and explicit morphosyntactic information than reasoning about world knowledge or pragmatic information
  d. the force of selection from child learning for high E-complexity being strong enough relative to drift to influence the typology of esoteric populations

e. the existence of a force of selection against high E-complexity
f. the hypothesis that this force originates in adult L2 learning
g. the force of selection for high E-complexity being weak enough relative to this second force of selection against high E-complexity to explain the typology of exoteric situations

In contrast, Burkett and Griffiths (2010) and Reali et al. (2014) suggest explanations that depend only on the existence of a force of selection against high E-complexity. Note that with respect to both Burkett and Griffiths (2010) and Reali et al. (2014), this force of selection is rooted in learning preferences that are uniform over all agents in all populations. Both suggest that effects of population size and structure – effects that would also be present under the LNH's assumptions – can create conditions in small, esoteric populations that are plausibly sufficient to allow hard-to-learn variants to be maintained there at a higher rate than in large, exoteric ones. In other words, both papers suggest an explanation strictly simpler than the LNH.

4 Conclusion

We have argued that scientific explanations of variation and change in evolutionary systems (including language change) are beset by two key problems: a lack of informative data and a wealth of logically possible explanations with unclear or plausibly overlapping predictions. Further, we have argued that responsible scientific investigation in the face of these problems requires clearly presented and preferably mathematically explicit models of hypothesized mechanisms (e.g., learning), as well as thorough consideration of neutral explanations, and that simpler explanations be preferred by default. Turning specifically to the LNH's adaptationist claim about the relationship between E-complexity and social situation, we have pointed out that both of the problems generally facing explanation in evolutionary systems are particularly acute for the LNH and that what mathematical models we do have suggest that there are simpler, neutral (or more neutral) explanations for why high E-complexity (or generally, a language variant that is selected against in general) would be expected to be found in smaller, esoteric communities (i.e., explanations that do not invoke or require there to be any selection for high E-complexity specifically in esoteric social situations). First, we discussed how small population size should be expected to *amplify* the role of a neutral process (evolutionary drift) and *mask* the effects of selection in shaping the state and trajectory of an esoteric community's language variety. Second, we have reviewed recent work on mathematical modeling of language change suggesting that learnability selection *against* a language variant (crucially without selection *for* it in any condition) could lead to its differential appearance and persistence in small, esoteric populations by causing greater homogeneity of input to learners compared to exoteric situations.

In sum, we conclude that, in the absence of compelling evidence that high E-complexity facilitates child learning or the presentation of specific evidence against neutral explanations for the relation between morphological typology and social situation, general principles of evolutionary systems and current models of language change suggest that the most likely explanation for the morphological typology of esoteric communities does not reflect adaptation to infant learning. While the LNH was partly intended to account for supposed correlations between what we have denominated the E-complexity of morphological systems in esoteric situations, it, correctly, does not assume that languages in such situations are either the only languages with high E-complexity or that they are necessarily more E-complex than those in exoteric situations. In fact, high E-complexity obtains for languages in very varied social situations, in many population sizes, and ranging over different areal distributions. For example, Hungarian, a member of the Ob-Ugric branch of the Uralic language family with 13,000,000 speakers, displays quite elaborate inventories of verbal and nominal marking; Mordvin, a member of the Volga-Finnic branch of the Uralic language family with approximately 400,000 speakers, possesses the most complex system of verbal inflection in Uralic; Navajo, a member of the Athapaskan family with approximately 145,000 speakers, contains an extraordinarily rich system of morphosyntactic and allomorphic variation in both its nominal and verbal systems (Bonami et al. 2019). From the perspective of parsimony, of course, we would like any account to cover the learning of all three languages as well as languages exhibiting even more complex and simpler systems. Given this, a real learnability condundrum remains and becomes plain: how does the learning of (complex) morphological systems actually occur, in both small communities and larger ones, esoteric and exoteric? We have hypothesized that this process is guided by morphological organization measured in terms of I-complexity, which is to say, patterns and subpatterns of conditional entropy that facilitate good guesses from known (patterns of) forms to unknown (patterns of) forms. Throughout we have alluded to connections between neutral theory, language change, and systemic morphological organization as synthesized in Lass (1997).[31] These connections are complex and must be explored elsewhere.

---

[31] See also Norde and Van de Velde (2016).