S-CURVES AND THE MECHANISMS OF PROPAGATION IN LANGUAGE CHANGE

RICHARD A. BLYTHE
University of Edinburgh

WILLIAM CROFT
University of New Mexico

A variety of mechanisms have been proposed in sociolinguistics for the propagation of an innovation through the speech community. The complexity of social systems makes it difficult to evaluate the different mechanisms empirically. We use the four-way typology of mechanisms proposed by Baxter and colleagues (2009), and define them mathematically in such a way that the effects of different mechanisms in the trajectory of a change can be modeled. The model suggests that the widely observed empirical pattern of an S-curve temporal trajectory of change can be captured only if the mechanisms for propagation include replicator selection, that is, differential weighting of the competing variants in a change, except under highly specialized circumstances that probably do not hold in speech communities in general.*

Keywords: S-curves, propagation, social network, mathematical modeling, evolutionary models, replicator, selection

1. INTRODUCTION. The mechanisms of propagation in language change, that is, how a novel linguistic variant comes to be adopted as the convention of a speech community, are still a matter of major debate in sociolinguistics. A great variety of mechanisms have been proposed, with empirical evidence being offered to confirm or disconfirm each of them. Yet much uncertainty remains, even among leading scholars.

For example, Labov, in his *Principles of linguistic change: Social factors* (2001), entertains two quite different mechanisms for the propagation of a change. On the one hand, he states that the position closest to his point of view is that of Sturtevant (1947), who 'viewed the process of linguistic change as the association of particular forms of speaking with the social traits of opposing social groups' (Labov 2001:24). Labov states that Sturtevant and LePage and Tabouret-Keller (1985), whose theory of acts of identity is similar to Sturtevant's, '[b]oth attribute linguistic change to: (1) the association of positively regarded traits and social privileges with membership in a given social group; and (2) the association of a linguistic form with membership in that social group' (Labov 2001:191).

On the other hand, Labov also suggests a different mechanism, namely the frequency of interaction of interlocutors, citing Bloomfield (1933; e.g. Labov 2001:19). Just after defining the Sturtevant-inspired theory, Labov writes, 'language change may simply reflect changes in interlocutor frequencies which are in turn the result of changes in social preferences and attitudes' (2001:191).

Labov does not rule out the second mechanism. But the two mechanisms differ radically in kind. In the Sturtevant-inspired mechanism, particular variants are associated with some type of social valuation, acquired by association with a social group or a certain class of individuals. In the Bloomfield-inspired mechanism, no difference in valua-

---

* Earlier versions of this work were presented at the Conference on 150 Years after Darwin: From Molecular Evolution to Language, Institute for Complex Systems, Palma de Mallorca, Spain; the Workshop on Language as an Evolutionary System: An Interdisciplinary Approach, Edinburgh, UK; the 43rd annual meeting of the Societas Linguistica Europaea, Vilnius, Lithuania; the Summer Institute of the Linguistic Society of America, Boulder, CO; and the 2011 European Conference on Complex Systems, Vienna, Austria. Blythe was supported by an Academic Fellowship from the Research Councils UK. The authors would like to thank Gareth Baxter, Alan McKane, Sali Tagliamonte, Peter Trudgill, and the audiences and students at the aforementioned presentations for their comments and input, and thank four anonymous referees, the associate editor, Harald Baayen, and the editor, Greg Carlson, for their comments. All errors remain our own.

tion is associated with particular variants; nor does any social status or attitude toward speakers or groups of speakers play any role in the propagation of a variant. Instead, frequency of interaction with interlocutors—higher with some, lower with others—is taken to drive the propagation of novel variants.

How many qualitatively distinct possible mechanisms of language change are there? And how can we decide among them, or at least determine which mechanisms are necessary and/or sufficient for explaining language change? Of course, empirical confirmation or disconfirmation is the most obvious and best way to answer these questions. But some questions are very difficult to address empirically. For instance, collecting data on interlocutor frequencies (the Bloomfield-inspired mechanism) is extremely difficult, and it is not clear whether we would be able to decide between interlocutor frequencies vs. associated social values as the best or only explanation for the propagation of a novel variant. The system of social behavior is so complex that empirical analysis can be very challenging.

There is another approach that offers the opportunity to address some—though certainly not all—of the theoretical issues of the mechanisms of language change in society, namely mathematical modeling of social systems and linguistic behavior. Mathematical modeling offers two significant advantages. First, it demands that we provide precise mathematical definitions of the proposed mechanisms that allow those mechanisms to be incorporated into a model. These definitions allow us to sharpen our theoretical tools for analyzing language change. We provide sharpened definitions for mechanisms of language change in §3. Second, the model of various mechanisms can be applied to empirically supported patterns of language change, and some mechanisms can be ruled out, or ruled in, based on the ability of the model to capture the empirical pattern with or without those mechanisms. We present results of this application in §§6–7.

There are relatively few applications of this approach to our knowledge (see §§2, 6). A typology of mechanisms of language change is offered by Baxter and colleagues (2009) in their model of Trudgill's theory of new-dialect formation. We provide a more precise definition of the mechanisms of language change identified by Baxter and colleagues, and discuss how various mechanisms proposed in the literature on the propagation of language change are categorized according to the typology. We then apply a model of these mechanisms to the empirically widely attested observation that the trajectory of a large class of language changes, namely replacement of a variant by a new competing variant, follows an S-curve. Our modeling indicates that language changes that follow an S-curve almost certainly involve at least differential valuation of linguistic variants. None of the other mechanisms, alone or in concert, produce an S-curve trajectory of language change except under highly specialized circumstances that are unlikely to hold in speech communities in general.

2. AN EVOLUTIONARY FRAMEWORK FOR LANGUAGE CHANGE AS A CONTEXT TO DEFINE MECHANISMS OF CHANGE.

Baxter and colleagues' (2009) model is based on the evolutionary framework for language change presented in Croft 2000, in turn based on Hull's GENERAL ANALYSIS OF SELECTION for evolutionary systems (Hull 1988, 2001; see also Dawkins 1989 [1976], Mufwene 2001, 2005, 2008). We briefly describe the evolutionary framework for language change in order to place the mechanisms of change in context.

Evolutionary systems are theories of change by replication, as opposed to inherent change (Hull 1988:410). Inherent change is internal change of a single entity, while change by replication is change in the structure of distinct entities that are produced by some sort of replication process. Language changes by replication in that linguistic structures are replicated in utterances: they are replicated from prior utterances produced by speakers or their interlocutors, mediated by the knowledge about their language that the speakers have acquired (Croft 2000:Ch. 2). Sociohistorical linguistics examines changes in the population of replicated structures—tokens of variants of linguistic variables—across speakers and across time.

Two important distinctions are made in the evolutionary framework that are relevant to the typology of mechanisms of language change. The first is that evolutionary change is a two-step process: the GENERATION OF VARIATION in the replication process, and the SELECTION of some variants over others (also called differential replication). The generation of variation is called innovation in language change, while the selection process is called propagation or diffusion. This distinction is not always clearly made in discussions of language change, because not all scholars treat it as a two-step process (Croft 2000:5). In this article, we limit our attention to propagation, that is, selection, and assume the existence of two or more competing variants (for discussion of mechanisms for the generation of variation, see Ohala 1989, Keller 1994, Croft 2000, 2010, Pierrehumbert 2001).

The second important distinction in an evolutionary framework is between the roles played in evolutionary change by a speaker and an utterance, or more precisely a token of linguistic structure in an utterance. Generalized models of evolutionary change such as Dawkins 1989 [1976] and Hull 1988 identify the entity that is replicated as the REPLICATOR. In biological evolution, the canonical example of a replicator is the gene. Croft's evolutionary model of language change posits the linguistic structures in utterances as the replicators in language change; as noted above, this conforms with sociohistorical linguistic approaches to language change, as well as usage-based models of language (Barlow & Kemmer 2000, Bybee 2001, 2007).

There is a second significant role in evolutionary change, called the INTERACTOR by Hull (and the 'vehicle' by Dawkins, who downplays its importance). According to Hull, the interactor is an entity that interacts with its environment to cause differential replication, that is, selection, to take place. Hull argues that in biological evolution, a variety of entities can function as interactors, including the gene itself. But the canonical example of an interactor is the organism: its interaction with its environment, determining its probability of survival and reproduction, causes replication of its genes to be differential. The canonical interactor in language change is the speaker: his or her choices of what to say and how to say it, by whatever mechanisms and for whatever reasons, determine which variants (replicators) are replicated or not.

In the basic evolutionary model of language change, speakers replicate linguistic structures in utterances while interacting with other speakers. Those tokens of linguistic structures are the replicators. The replication process generates variation (produces innovation), via mechanisms that will not be examined here. Once these variants are available to speakers, speakers choose—not necessarily consciously or intentionally—to produce certain variants. Mechanisms of linguistic selection lead to the differential replication, that is, propagation, of some variants at the expense of others.

The general analysis of selection does not specify the mechanisms by which the generation of variation and selection happen in a specific domain. In biology, variation is generated by random mutation; selection occurs through functional adaptation to the environment, or change occurs through genetic drift. The domain-specific mechanisms of biology will not apply to language. The more abstract types of evolutionary processes—generation of variation, selection, and drift—apply to language, however. In fact, as demonstrated in the next section, Hull's general analysis of selection implies two further mechanisms of selection in addition to classical selection and drift. In the next section, these four mechanisms, introduced by Baxter and colleagues (2009), are given formal, model-independent analyses, and are related to theories of language change in sociolinguistics.

3. A TYPOLOGY AND ANALYSIS OF MECHANISMS OF PROPAGATION IN LANGUAGE CHANGE.

The two mechanisms illustrated in §1—the Sturtevant-inspired and Bloomfield-inspired mechanisms—are two of four qualitatively different mechanisms of selection identified in Baxter et al. 2009:267–72. Baxter and colleagues do not discuss the sociolinguistic theories that can be identified with these mechanisms, apart from the theory of Trudgill's that they model. We introduce here more formal definitions of these mechanisms that are independent of any specific model implementing them. Our formal analysis is presented in prose in this section, and through equivalent mathematical formulae in the appendix.

The input to propagation is a range of variants generated by the first step of evolutionary change. The processes that generate linguistic variation produce many variants, including near-continuous variation in some cases (Pierrehumbert 2001, Croft 2010). We follow general sociolinguistic practice in dividing potentially continuous variation into a number of discrete variants of a linguistic variable. The results presented here assume two variants—the original convention and the variant that may replace it. Extensions to the case of multiple variants, however, could be included within the scheme we describe.

3.1. REPLICATOR SELECTION.

The Sturtevant-inspired mechanism is characterized by differential valuation of the specific variants. Because the variants are the replicators in language change, Baxter and colleagues describe this as REPLICATOR SELECTION. We introduce a formal definition through the following thought experiment. Suppose two variants, A and B, are present with some prescribed distribution of usage frequencies across members of a speech community at some initial time. After some number of interactions between speakers, these frequencies will change, and there will be some probability distribution of usage frequencies of the two variants at that time. Now consider a parallel speech community, identical in every way to the first, other than that the initial frequencies of the two variants are reversed. We can now compare the distributions of usage frequencies at the same point in time between the two communities. If after exchanging the usage frequencies of the two variants in the first, there is the same probability distribution of usage frequencies in the second community, then there is no differential valuation of one variant over the other. The variants are replicated with the same frequency when used in the same social and linguistic contexts: that is, they are symmetrical. The signature of replicator selection is an absence of this symmetry.

The mechanisms proposed in Labov's classic research are instances of replicator selection: linguistic forms are associated directly with social values, and only indirectly with the individuals or social groups that lent the forms their social value (see §7.3). Various social factors have been proposed to lead to differential social values for variants. Labov posits prestige, defined in socioeconomic terms, as the typical factor for changes from above (1972:290, 1994:78; but cf. Labov 2007:346, n. 2). For changes from below, Labov first proposed a concept of 'covert prestige' (Labov 1966:108; see also Trudgill 1972); more recently, Labov has argued that the variants' weighting comes from nonconforming but socially mobile individuals (Labov 2001, especially pp. 511–18). LePage and Tabouret-Keller (1985) propose that the weighting of variants derives from the acts of identity with a group that they represent. Mufwene describes competition among variants as 'the unequal ways the variants and coexistent language varieties are weighted by their speakers' (Mufwene 2008:4).

The replicator selection mechanisms for propagation proposed by sociolinguists are all social. Other linguists have proposed other mechanisms for propagation, including phonetic biases and morphological analogy favoring some variants over others for sound change, and structural or functional biases favoring some variants over others for grammatical and lexical change (e.g. Haspelmath 1999, Nettle 1999). These types of explanations would also be instances of replicator selection, since one variant would be weighted differentially over another. Croft argues, however, that these phonetic, structural, and/or functional biases are mechanisms for the generation of variation (in contrast to biological evolution, in which the generation of variation appears to be random; Croft 2000:8).

Our mathematical model cannot decide among these various social, functional, structural, or phonetic weightings of variants, or indeed any other differential weighting of variants (on the use of prestige in sociolinguistics, see Labov 2001:24 and Trudgill 2004:153; on the use of acts of identity, see Trudgill 2008). But all of these theories are instances of replicator selection.

3.2. NEUTRAL INTERACTOR SELECTION. The other mechanism described in the introduction, the Bloomfield-inspired mechanism, operates in a very different way. Language change is simply driven by the frequency of interaction with one's interlocutors. As we observed in §1, no social valuation of the variants used plays a role, nor does any social valuation of the speaker using the variant(s). All that matters is how often one converses with one interlocutor as opposed to another, and which variants those interlocutors produce in the conversational interaction. For this mechanism, the variants are symmetrical: if one replaced variant A with variant B in the manner described above, then variant B would have the same path of propagation as variant A: all that matters is who is talking to whom.

This mechanism exhibits a symmetry between pairs of speakers that defines what Baxter and colleagues (2009) call NEUTRAL INTERACTOR SELECTION. We formalize the definition as follows. Suppose that a speaker labeled i is the sole user of the innovative variant (say, B) at some initial time. After some number of interactions, a second speaker j has some probability of using the variant with a prescribed frequency x. As previously, we consider a parallel speech community, which differs from the original one in that speaker j is the sole user of the innovation instead of speaker i. The symmetry defining neutral interactor selection is present if, at all points in time, the probability that speaker i in the second community is using the innovation with frequency x is equal to the corresponding probability for speaker j in the first community. Another way to express this is to say that a variant's path of propagation from speaker i to speaker j has the same probability of occurring as the reverse path from j to i.

The terminology 'neutral interactor selection' emphasizes that it is the properties of the interactors, the interlocutors, that drive the diffusion process rather than properties of the replicators. These interactor properties do not include any sort of social valuation of the interactor/speaker. Hence it is neutral in the evolutionary sense of that term (on which more below). Nevertheless, it is an instance of selection as defined by Hull. In Hull's general analysis of selection, selection is a process by which the interactor's interaction with its environment causes differential replication of the replicators.[1] Part of the interactor's environment is the other interactors. In the case of language change, the other interactors are the interlocutors. In this mechanism, it is the interlocutor's fre-

---

[1] Actually, Hull's definitions (1988:408–9) of interactor and selection differ slightly from each other. He defines an interactor as 'an entity that interacts as a cohesive whole with its environment in such a way that

quency of interaction with other interlocutors that causes propagation, that is, differential replication of linguistic variants. Hence the process is a type of selection process.

It is also possible for frequency-based biases to operate while respecting neutral interactor selection. For example, a speaker may boost the production frequency of the variant she uses most often, or of a variant that has recently been used by an interlocutor (possibly through accommodation or priming effects AS LONG AS these biases are applied independently of the identity of an interlocutor). Furthermore, if replicator selection is absent, these biases must also be applied in the same way to all variants. We remark that these typically nonlinear frequency biases have an analog in biological evolution as frequency-dependent selection, one possible consequence of which is the stable coexistence of multiple variants (see e.g. Ayala & Campbell 1974 and Nowak 2006 for reviews).

Trudgill's deterministic theory of the emergence of isolated new dialects such as New Zealand English from dialect mixture is a clear example of neutral interactor selection (Trudgill 2004; Peter Trudgill, p.c.). The basis for Trudgill's theory is the theory of accommodation (Giles 1973, Giles & Smith 1979, Trudgill 1986:2): human beings tend to converge in their behaviors, such as linguistic behaviors, in social interaction, possibly for biological reasons (Trudgill 2004:28, 2008:252). Another example of neutral interactor selection is the social network theory associated with Milroy and Milroy (e.g. L. Milroy 1987, Milroy & Milroy 1985, 1992, J. Milroy 1992). Social network structure includes differences in tie strength between speakers. Tie strength is more than interlocutor frequencies. It shares three features with simple interlocutor frequencies, however. First, there is no differential valuation of variants. Second, it is a function only of which persons a speaker interacts with, including the strength of the interaction. Finally, tie strength is a symmetric function: if speaker A has a strong tie with speaker B, then speaker B has a strong tie with speaker A—the same frequency and multiplexity of interaction. Thus, in neutral interactor selection, there is symmetry in the behavior of variants (replicators) and the behavior of speakers (interactors). Both the Milroys and Trudgill argue that weak ties between speakers function as conduits for novel variants to spread throughout a speech community (Milroy & Milroy 1985:365, Trudgill 2004:162; see also Labov 2001:364).

3.3. WEIGHTED INTERACTOR SELECTION. It is of course possible for a speaker's linguistic behavior to be more strongly affected by one speaker over another, and in particular, for a speaker *i* to have a greater influence on another speaker *j* than *j* has over *i*. In this case, the symmetry defining neutral interactor selection no longer applies. Baxter and colleagues (2009) call this WEIGHTED INTERACTOR SELECTION. Weighted interactor selection is not neutral between interactors (speakers, in language change); some are more influential than others.

Weighted interactor selection can be identified in another model of propagation discussed by Milroy and Milroy (1985): the model of diffusion of innovations by Rogers (1995; Milroy & Milroy 1985:367). Rogers's model covers a range of mechanisms, but its distinctive feature, and the one discussed by Milroy and Milroy, is the partition of

---

this interaction CAUSES replication to be differential', and selection slightly differently as 'a process in which the differential extinction and proliferation of interactors CAUSES the differential propagation of the relevant replicators' (emphasis original). We follow Hull's definition of interactor in defining selection, which is in accordance with Hull's intention; the crucial feature is the causal relationship from the interactor's interactions to the differential replication of the replicators (David Hull, p.c.).

members of a community in terms of their receptiveness to adopting an innovation. Rogers's model partitions the community into the following five ADOPTER CATEGORIES (Rogers 1995:261–66).

(i) Innovators: those who are most receptive to innovations and will transmit them to other members of the community
(ii) Early adopters: opinion leaders in the community who are respected by the rest of the community
(iii) Early majority: those who willingly adopt an innovation, but only after early adopters have done so
(iv) Late majority: those who wait until a majority of members of the community have adopted an innovation before adopting it themselves
(v) Laggards: those who resist innovations, and 'possess almost no opinion leadership'

In the sociolinguistics literature, most attention has been devoted to innovators and early adopters, in effect leading to at most a three-way partition of speakers (innovators, early adopters, later adopters/laggards). Labov (2001:356–60) uses a two-way partition into leaders and followers, following Katz and Lazarsfeld (1955; see also Rogers 1995:284–86); see §6.3 for further discussion.

What all these models have in common, however, is that the relationship between members of different adopter categories is not symmetrical. For example, later adopter or follower categories follow early adopters, and therefore weight the behavior of early adopters more than the behavior of their own cohort; but early adopters/leaders do not follow later adopters/followers—that is, they do not give a greater weight to the later adopters/followers' behavior than to their own cohort. This is what marks it out as an instance of weighted interactor selection.

Weighted interactor selection differs from replicator selection. In weighted interactor selection, what matters is the social group that uses certain variants, not the variants themselves. The forms used by a particular group, for example, the early adopters/leaders, are favored because they are used by that group; if the same form happens to be produced by a member of another group, say the laggards, it will not have the same weighting. In weighted interactor selection the social valuation is attached to the speaker producing the variant, not the variant itself. In replicator selection, by contrast, the social valuation is attached to the form, no matter who produces that form, even if the form originally gained that social valuation by virtue of being associated with a particular social group (see §7.3).

3.4. NEUTRAL EVOLUTION (DRIFT). Finally, Labov suggests that some sound changes occur without any social mechanisms operating at all: a theory invoking a social mechanism such as his own 'does not apply to most mergers and other changes which run their course without taking on symbolic value of any kind' (Labov 2001:517). What Labov is suggesting here is that language change can happen without any social process being involved. This is an example of invoking the last mechanism described by Baxter and colleagues—NEUTRAL EVOLUTION. Neutral evolution, also known as 'drift' in biological evolution (see §2), is change that results from the random fluctuations in replicator frequencies in a finite population: these fluctuations may go to 0%, leading to extinction of the replicator, or 100%, leading to fixation of the replicator. The evolutionary concept of drift in biology is not the same as linguistic drift (Sapir 1921), in which the seemingly autonomous change is directional, not random, particularly in the case in which daughter languages change in parallel but not via contact (see also

Trudgill 2004:129–47). For this reason, we follow Baxter and colleagues in using the term 'neutral evolution' rather than 'drift' for this mechanism.

Neutral evolution is related to neutral interactor selection in that neither mechanism involves any differential weighting of either interactors (speakers) or replicators (variants). Not surprisingly, neutral evolution as a mechanism for language change has not been examined in detail by sociolinguists, as it has no social dimension to it. Neutral evolution's independent role in evolutionary change can only be inferred in models with a 'flat' network structure, in which all individuals have an equivalent position within the social structure. For example, models in which speakers have equal probability of interacting with all other individuals (i.e. all individuals are directly connected, and there are no differences in tie strength between individuals) fall into this category. Examining such models mathematically demonstrates that the likelihood of fixation (propagation to 100%) of a replicator (e.g. linguistic variant) is proportional to its frequency in the population. Hence Trudgill's observation that the majority variant in the input to New Zealand English is generally the one that became fixed for the New Zealand English dialect (Trudgill 2004:113–15) is probably a consequence of the operation of neutral evolution rather than neutral interactor selection.

3.5. COMPARISON AND COMBINATION OF THE MECHANISMS. In our analysis of the mechanisms of propagation of change, the salient properties of the different mechanisms are whether their effects on replicators and interactors are symmetric or not. In replicator selection, the replicators are nonsymmetric (i.e. not necessarily symmetric, though for replicators of equal weighting, of course, they will be interchangeable). In weighted interactor selection, replicator behavior is symmetric, but the interactors (speakers) are nonsymmetric: speaker A may weight the language produced by speaker B more heavily, but not vice versa. In neutral interactor selection, replicator behavior and interactor behavior are symmetric; all that matters is interaction frequency, which may be unequal (speaker A may interact more with speaker B than with speaker C). Finally, neutral evolution is the behavior found with no interaction inequalities, and symmetric interactor and replicator behaviors. The major properties distinguishing the four mechanisms of propagation of a change are summarized in Table 1.

	INTERACTION FREQUENCY	INTERACTOR BEHAVIOR	REPLICATOR BEHAVIOR
NEUTRAL EVOLUTION	equal	symmetric	symmetric
NEUTRAL INTERACTOR SELECTION	unequal	symmetric	symmetric
WEIGHTED INTERACTOR SELECTION	—	nonsymmetric	symmetric
REPLICATOR SELECTION	—	—	nonsymmetric

TABLE 1. Typology of the mechanisms of propagation of language change.

The four mechanisms can be ranked by 'strength' or 'complexity'. Neutral evolution evokes nothing more than random processes. Neutral interactor selection adds differential rates or levels of interaction among interactors. Weighted interactor selection adds (possibly asymmetric) weighting of the interactors producing the replicators (variants). Finally, replicator selection is the only mechanism in which the replicators themselves are differentially weighted. This ranking is recognized by sociolinguists who compare the mechanisms. For example, Labov writes, 'The account based on covert attitudes [manifested in the social valuation of variants, i.e. replicator selection] is redundant to the extent that the network daily interaction brings people into contact with the new form in proportion to their distance from the originating group [i.e. neutral interactor selection]' (2001:192). In other words, Labov suggests that if the weaker mechanism, neutral interactor selection, is sufficient to account for the data, then it should be adopted over the stronger mechanism, replicator selection.

Theories of language change may and do involve multiple mechanisms of more than one type, though there is sometimes unclarity about what mechanisms are involved. Mufwene relies primarily on replicator selection, as described above. At the opposite end of the spectrum, Trudgill relies exclusively on neutral interactor selection (and neutral evolution) for his theory of new-dialect formation. Labov's theory and the Milroys' theory invoke multiple mechanisms. As noted in §1, Labov also entertains the idea that neutral interactor selection might be involved in language change, although he does not incorporate it into his final theory.

Milroy and Milroy place neutral interactor selection as the centerpiece of their theory of language change. They introduce Rogers's model to sociolinguistics, which makes it appear that they are also invoking weighted interactor selection. But Milroy and Milroy (1985:369) criticize Labov for focusing on the prestige of the innovators, 'paying less attention to the content or structure of INTERPERSONAL LINKS' (emphasis original). In other words, they do not interpret Rogers's model as requiring weighted interactor selection. The use of Rogers's model by other sociolinguists such as Sankoff and Blondeau (2007; see §6.3), however, may reasonably be interpreted as invoking weighted interactor selection.

Milroy and Milroy go on to state that 'although a successful innovation needs in some sense to be positively evaluated, generalizations can be made about the social mechanisms controlling innovation and diffusion quite independently of the prestige value attached to any given innovation' (Milroy & Milroy 1985:369–70). Here, Milroy and Milroy acknowledge a role for replicator selection (a positive value associated with innovative variants). Even so, they consider the operation of replicator selection to be less important than network structure, that is, neutral interactor selection.

We conclude this section by remarking that although we assume the 'usage-based' theory of language change found in sociohistorical linguistics in this article, the typology of selection mechanisms can also be applied to other types of models, for example, 'child-based' models in which a language changes through learners failing to acquire the exact same grammar used by adult members of the community (but see Croft 2000 and Bybee 2010 for critiques of the child-based theory). These are also evolutionary models: replication is the grammar acquisition process; variation is generated by allowing children to acquire grammars that are not identical to their parents' grammars; and there is differential replication of different grammars across speaker populations (Niyogi & Berwick 1997:698, Croft 2000:44). In Niyogi and Berwick's model, sustained directed change, where it occurs, is a consequence of replicator selection. Note that this identification of a selection mechanism contrasts with the analysis of Briscoe (2000:249), who points out that the induction algorithm is neutral (i.e. if one cannot decide from the data which grammar is more likely, the argument is resolved by a coin toss). By using the definition of replicator selection in terms of symmetries, however, we see that the structure of the grammar combined with the induction algorithm constitutes a form of replicator selection.

4. APPLYING THE MODEL: S-CURVES IN THE TRAJECTORY OF LANGUAGE CHANGE.

The type of language change whose temporal trajectory we analyze here is the replacement of a competing variant of a linguistic variable. Language changes can be divided into two broad types: changes in the linguistic system and changes to a linguistic variable.

SYSTEMIC changes include the open-ended increase of vocabulary in the languages used in modern industrial society, or the change from an inflectional language to an isolating one by the loss of all of the language's inflections. Systemic changes are the cumulative effect of individual variable changes. These cumulative effects do not necessarily follow the same principles as changes to individual variables. For example, there does not appear to be a tailing off to a ceiling value in the aforementioned increase in vocabulary (see e.g. Meibauer et al. 2004, Scherer 2007).

VARIABLE changes may be divided into types based on origin and outcome. A new variant may INTRODUCE a new linguistic convention, such as a new word for a new artifact, or the grammaticalization of a plural marker in a language without number marking. Or a new variant may COMPETE with an existing variant used for an established convention. We are not aware of quantitative studies of the trajectory for introduced variants. Finally, in the case of competing variants, a new variant may end up REPLACING an existing variant, or REALLOCATION may take place—that is, the two variants will become specialized in meaning, register, or both, and therefore cease to be in competition (Trudgill 1986:110, Croft 2000:176). In the latter case, the variable splits in two, and each variant goes to (near) completion in its own function. Reallocation involves a change in the variable in the middle of its trajectory. Also, a variable change may be interrupted or halt in the middle of its trajectory for reasons other than reallocation. For example, in the rise of periphrastic *do* in English, periphrastic *do* halted its rise in positive declarative contexts, while it followed an S-curve to completion in negative and interrogative contexts (Ellegård 1953, Nevalainen & Raumolin-Brunberg 2003:68–70).

The best-documented cases of the trajectory of language change are instances of replacement of a conventional variant by a competing variant in a linguistic variable. Our analysis must therefore be restricted to that type of language change, although it is reasonable to hypothesize that introduced variable changes have a similar trajectory, and interrupted changes follow a similar trajectory before the interruption.

It has long been asserted that replacement changes follow a trajectory that is best approximated by an S-curve (e.g. Greenberg et al. 1954:155, Weinreich et al. 1968: 113, Bailey 1973:77, Kroch 1989:203, Labov 1994:65, Ogura & Wang 1996:119 (for lexical diffusion), Croft 2000:183, Chambers 2002:361, Denison 2003, Nevalainen & Raumolin-Brunberg 2003:53). Chambers writes that '[t]he S-curve has since been observed in diffusions of all kinds … , and is now established as a kind of template for change' (2002:361).

A survey of thirty-nine changes documented in the sociolinguistics literature indicates that this assertion is a plausible one (see Table 2). Twenty-two of the changes, documented in both real time and apparent time,[2] can reasonably be fit to a full S-curve: the

---

[2] In apparent-time studies, changes in the frequencies of variants are observed across age groups in a synchronic sample of speakers. Apparent-time studies are hypothesized to reflect real-time changes on the assumption that language behavior is relatively fixed after adolescence. Recently, evidence has been presented that adults do adjust their production of variant frequencies (e.g. Harrington et al. 2000, Sankoff & Blondeau 2007). The consensus, however, is that apparent time is still a useful construct (see Bailey 2002, Tagliamonte & D'Arcy 2009:61–63, and references cited therein): older speakers shift in the same direction as younger speakers, therefore reinforcing the directionality of the change; and older speakers are still much less labile than younger speakers.

The study by Tagliamonte and D'Arcy in Table 2 reveals another deviation from the S-curve that is frequently found in apparent-time studies: the 'adolescent peak' (Labov 2001, Tagliamonte & D'Arcy 2009). In many apparent-time studies, the frequency of the incoming variant rises in an S-shaped curve from oldest to younger speakers, but peaks with adolescent speakers and then declines for preadolescents. This perturbation of the S-curve is a consequence of the process of language acquisition and intergenerational transmission, and does not affect our overall conclusions about the mechanisms of propagation that yield the S-shaped trajectory.

observed data points, ranging up to a dozen points for a single change, consistently show a slow beginning, a rapid uptake, and a slow tailing off for the propagation of the innovation. In addition, Labov suggests that a systemic change also follows an S-curve. Thirteen changes may be interpreted as the beginning or the end of an S-curve trajectory. Of course, this analysis is not necessary for these changes, but it is compatible with the S-curve hypothesis, given the absence of data in a longer time interval that would include the full replacement process. Finally, there are three changes in our survey that appear to stop and go in reverse. These may be interpreted as changes following an S-curve trajectory that are then interrupted; we do not analyze such changes here.

LANGUAGE	CHANGE	VARIANTS	TIME	CURVE	SOURCE
Brazilian Portuguese	future tense	4	R	F	Poplack & Malvar 2006, 2007
French	negation	4	R	F	Grieve-Smith 2009:143, table 10
Russian	zero genitive in units of measurement	2	R	F	Altmann et al. 1983
Chinese, Shuāng-fēng dialect	lexical diffusion of voicing split of initial obstruents	2	R	F	Wang & Cheng 1970
English	periphrastic *do,* negative and/or interrogative contexts	2	R	F	Ellegård 1953, Kroch 1989
English	negation	3	R	F	Grieve-Smith 2009:27, data from Mazzon 2004:25, 56
English	generalization of subject *you*	2	R	F	Nevalainen & Raumolin-Brunberg 2003:60
English	loss of multiple negation	2	R	F	Nevalainen 2000:354–56 (1460 to 1681); Nevalainen & Raumolin-Brunberg 2003:71–72 (only to 1619)
English	generalization of *-(e)s*	2	R	F	Nevalainen & Raumolin-Brunberg 2003:68
English	object of gerund, zero vs. *of*	2	R	F	Nevalainen & Raumolin-Brunberg 2003:66
English	inversion after negators	2	R	F	Nevalainen & Raumolin-Brunberg 2003:73
English	generalization of *are*	2	R	F	Nevalainen 2000:359
Scots English	replacement of indefinite article *ane* by *a/an*	2	R	F	Devitt 1989:23–26, 36–37
English (East Anglia, East Midlands)	lexical diffusion of changes in (u)	2	A	F	Chambers & Trudgill 1998:161–64
Toronto English	quotative forms	5	A	F	Tagliamonte & D'Arcy 2007:205, figure 2.5
Canadian English	terms for 'sofa'	3 major, 8 minor	A	F	Chambers 1995
Canadian English	seven changes, lexical and phonological (incl. 'sofa')	2	A	F	Chambers 2002:366, 369, data in Chambers 1998
Philadelphia English	system of vowels	systemic	A	F	Labov 1994:67
English	periphrastic *do*, negative contexts	2	R	B	Nevalainen & Raumolin-Brunberg 2003:70
English	prop-word *one*	2	R	B	Nevalainen & Raumolin-Brunberg 2003:64

(TABLE 2. *Continues*)

LANGUAGE	CHANGE	VARIANTS	TIME	CURVE	SOURCE
English	noun subject or gerund	3	R	B	Nevalainen & Raumolin-Brunberg 2003:67
English	possessive determiner *its*	3	R	B	Nevalainen & Raumolin-Brunberg 2003:62–63
English	indefinite pronouns with singular human reference	4	R	B	Nevalainen & Raumolin-Brunberg 2003:77–78
French	inversion with preposed direct objects	2	R	B	Kroch 1989:215, data from Priestley 1955:19
Scots English	replacement of *quh-* relative clause markers by *wh-*	2	R	B	Devitt 1989:18–21, 36 (orthographic change)
Scots English	replacement of preterite *-it* by *-ed*	2	R	B	Devitt 1989:21–23, 36
English	*my/thy* vs. *mine/thine*	2	R	E	Nevalainen & Raumolin-Brunberg 2003:62
English	loss of inversion after initial adverbs	2	R	E	Nevalainen & Raumolin-Brunberg 2003:73
English	*which* vs. *the which*	2	R	E	Nevalainen & Raumolin-Brunberg 2003:74
Scots English	replacement of negative *na/nocht* by *no/not*	2	R	E	Devitt 1989:27–28, 37 (orthographic change)
Scots English	replacement of present participle *-and* by *-ing*	2	R	E	Devitt 1989:28–30, 37
English	periphrastic *do*, positive declarative contexts	2	R	I	Kroch 1989:224; Nevalainen & Raumolin-Brunberg 2003:68–70
English	replacement of *-'s* genitive	2	R	I	Rosenbach et al. 2000
English	prepositional phrase vs. relative adverb (vs. stranding)	3	R	I	Nevalainen & Raumolin-Brunberg 2003:75–76

Time: R: real, A: apparent
Curve: F: full, B: beginning, E: end, I: interrupted

TABLE 2. Survey of variable replacements documented in sociolinguistics literature.

To our knowledge there are no clearly documented cases of a change going toward completion that follows either a simple linear trajectory or an exponential curve (either slow start with a rapid completion and no tapering off, or an immediate rapid increase followed by a slow completion rate). There are, however, examples of variation that does not seem to be going toward completion, at least in the documented time period. These examples appear to exemplify either reasonably stable variation with the variants fluctuating around a mean percentage value, or a rise and fall of a relatively low-frequency variant, commonly in competition with an incoming variant that is going to (near) completion on an S-shaped trajectory.

One striking feature in the replacement by a competing variant is that the aggregate proportion of the incoming variant has an S-shaped trajectory despite variable behavior for subclasses of the change. For example, Poplack and Malvar (2006, 2007) argue that the multiple variants of the Brazilian Portuguese future entered via restricted contexts, and over the trajectory of the change, the favorable contexts for each variant changed as well. Yet the overall shape of the incoming *ir* 'go' periphrasis is an S-curve, as seen in Figure 1 (the other variants—the original synthetic future, the *haver*-periphrasis, and the futurate use of the present—rise and fall).

Devitt also observes an S-shaped curve for the anglicizing changes in Scots English in texts dating from 1600–1619 (Devitt 1989). Devitt observes that many texts in the time period have either a very low or very high frequency of anglicizing forms, while a

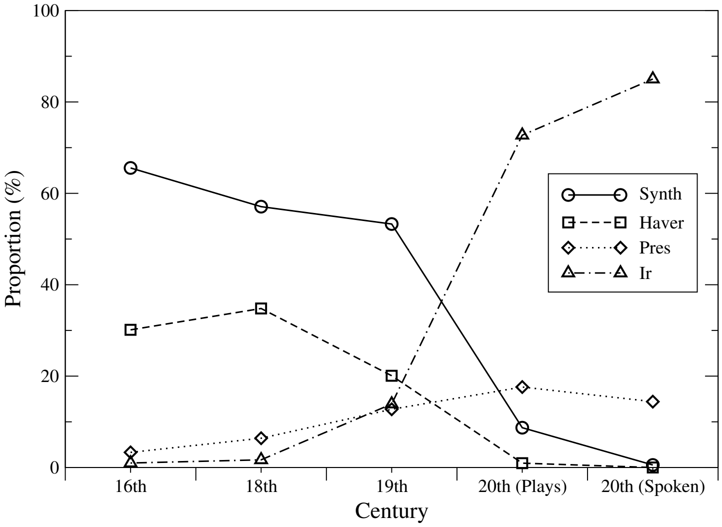

FIGURE 1. Trajectory of the evolution of four variants of the future in Brazilian Portuguese. Although three variants compete with the original synthetic future, the incoming *ir* 'go' periphrastic future is propagated following an S-curve. Data from Poplack & Malvar 2007:144.

few texts are found with frequencies in the intermediate range (Devitt 1989:38–46). Chambers observes that the merger of /w/ and /hw/ (in apparent time) takes a slightly different trajectory in different regions of Canada, with some speeding up later than others, although all appear to begin and end at approximately the same point (Chambers 2002:360–64). Nevalainen notes regional/social differences in the real-time changes in the rise of Standard English that she observes (Nevalainen 2000:347–56). In sum, the overall changes described display an S-shaped curve despite the variation in the behavior of individual words, speakers, texts, geographical regions, or social classes over the trajectory of the change.

5. COMPARING THE MECHANISMS TO PRIOR MATHEMATICAL MODELS OF SOCIAL DIFFUSION. Our aim in the rest of this article is to understand how the different selection mechanisms defined in §3 are related to the trajectory of propagation of an innovation. In particular, we want to determine which selection mechanisms lead to an S-curve, and which do not. A partial answer to this question can be obtained through a brief survey of existing agent-based models of social diffusion, which is presented in this section. None of these models, however, are set up to easily allow the different mechanisms to be enabled or disabled independently, thereby preventing a systematic investigation. This we provide in §6.

An important early agent-based model is due to Nettle (1999). It is rather complicated, but allows all the selection mechanisms of §3 to be implemented through specific parameter choices. In particular it allows for what Nettle calls 'social' and 'functional' selection, which relate to weighted interactor and replicator selection respectively. Nettle is most concerned with the THRESHOLD PROBLEM—the difficulty an innovation faces in becoming the new convention due to its necessarily low frequency when it has just been innovated. Nettle finds that an innovation can propagate with either functional selection or a high degree of social selection. The shape of the change trajectory, however, is not discussed in detail with reference to these mechanisms.

Nettle also reports on other outcomes, for example, stable coexistence of two variants when the probability that a speaker adopts an innovation is a nonlinear function of the number of interlocutors using it. These outcomes have been widely observed in models discussed in the mathematics and physics literature (see Castellano et al. 2009 for a comprehensive review of this vast array of models). The baseline model for social diffusion in this literature is the VOTER MODEL (Clifford & Sudbury 1973), which is an explicit instance of neutral evolution (see §3).

The voter model has a key property that is important in the following: if an innovation is present at frequency x at some time, the probability that it is conventionalized is equal to x. If an innovation initially has a very low frequency, it will most likely disappear rapidly before spreading through the community. That is, most innovations fail to propagate, as is most likely the case in reality. We are interested in those cases where the innovation does propagate. It is widely understood that in the voter model and its relatives, fixation is arrived at through fluctuations alone (e.g. Crow & Kimura 1970, Castellano et al. 2009). This behavior does not depend on whether usage of a linguistic variant is categorical (as in the voter model) or variable (as in the models of Baxter et al. 2006 or Reali & Griffiths 2010). As is demonstrated explicitly in §6, the change trajectories under neutral evolution are undirected and exhibit large fluctuations on the path to fixation: they are not S-curves.

This contrasts with a recent finding of Reali and Griffiths (2010), who show that neutral evolution is the consequence of a specific Bayesian learning algorithm. They argue that an S-curve is indeed obtained by conditioning on a change going to completion. In that work, however, the S-curve is obtained by averaging multiple change trajectories between the same initial and final state. Each of the documented instances of language change (see §4) provides a single instance of a change trajectory, each of which is an S-curve without performing any such averaging. Therefore, in the next section we focus on models and mechanisms that give S-curves in individual simulation runs.

Other agent-based models differ from the voter model in two main respects. Some models introduce nonlinear relationships between an innovation's usage frequency and its adoption probability. For example, these may boost the production frequency of the majority variant (as described in §3). Alternatively, frequency-dependent nonlinearities could be used to model theories due to Ogura and Wang (1996) and Grieve-Smith (2009) in which high token frequency verbs experience changes that start earlier, but proceed more slowly, than low token frequency verbs. In a biological context, such nonlinearities are referred to as frequency-dependent selection (Ayala & Campbell 1974, Nowak 2006). In a social context, it is typically found that majority-boosting rules dramatically reduce the probability that an innovation propagates, due to its initially tiny usage frequency (see e.g. de Oliveira 1992). Conversely, a bias toward the lower-frequency variant tends to lead to stable coexistence of both, since the variant in the majority is suppressed in favor of that in the minority. In population genetics this phenomenon has been referred to as stabilizing (or balancing) selection (see e.g. Hartl & Clark 2007); there are a number of socially inspired models in which this also occurs (Castellano et al. 2009:604).

The second way in which models depart from the voter model is by introducing an asymmetry between linguistic variants (i.e. replicator selection as defined in §3). This is seen in various models of language change—for example, Pierrehumbert 2001, Abrams & Strogatz 2003, and Ke et al. 2008. Although these models have been introduced in different contexts (for example, phonological change within a single speech community and competition between languages through language contact) and have distinct implementations, one common finding emerges: when there is sufficient replicator selection (referred to variously as ‘functional bias’ or ‘status’ of a language in these studies) to overcome any frequency-based biases (over- or underproduction of the majority), a sustained, directional change toward adoption of the innovation is observed. Again, however, the shape of the trajectory is not always reported.

The models surveyed here show a number of differing patterns of propagation of changes. In particular, there is some evidence that the characteristic asymmetry of replicator selection may tend to lead to a sustained, directed change from one convention to another. Since these patterns are based on models that differ in a wide variety of ways, however, it is not clear that any difference in a change trajectory between models is purely due to a different fundamental selection mechanism that is operating. We reiterate that what is needed to demonstrate the connection is a single, unified model that allows each mechanism to be activated independently. This further allows interactions between them to be investigated systematically.

6. RESULTS: PROPERTIES OF THE TRAJECTORY OF CHANGE ACROSS THE MECHANISMS. The unified model we introduce to scrutinize the relationship between selection mechanisms and change trajectories is an extension of the UTTERANCE SELECTION MODEL (USM) of Baxter et al. 2006, 2009. A complete technical definition of this new model is given in the mathematical appendix to this article. Here we restrict our discussion to the features that are most pertinent to our investigation of S-curve trajectories. Since the model is conceptually very similar to the USM, a reader who is unfamiliar with the use of agent-based models to describe language change may benefit from reading this section in conjunction with the more expansive account of the USM in a linguistic context provided in Baxter et al. 2009.

In the USM, speakers are characterized by the relative frequency (i.e. fraction of the time) that they use an innovation (as opposed to an existing convention). The value of this frequency for speaker *i* is denoted by the symbol x_i and lies between zero and one. The language—as specified by the entire set of frequencies x_i—changes over time as a consequence of interactions (conversations) between pairs of speakers.

What happens in an interaction is as follows. First, two speakers, labeled *i* and *j*, are chosen to speak to each other. Different pairs of speakers need not be picked with the same probability, reflecting the fact that speakers in a community do not all interact with each other with the same frequency. We use the symbol G_{ij} to denote the probability that the specific pair (i,j) is chosen to interact in any one timestep.

Then, both speakers produce an utterance. Each is assumed to contain T tokens of the linguistic variable of interest. For speaker *i*, an uttered token is of the innovative variant with probability x_i and of the conventional variant with probability $1 - x_i$. Likewise, speaker *j* produces these variants with probability x_j and $1 - x_j$, respectively. Thus on average, speaker *i* produces $n_i = x_iT$ innovative tokens, and speaker *j*, $n_j = x_jT$ innovative tokens. Because each token production is an independent random event (like the toss of a biased coin), however, the actual token counts n_i and n_j are random variables and vary from interaction to interaction even if x_i and x_j are unchanged.

After the tokens have been produced, the speakers combine the two token counts to form a single empirical frequency of the innovation. This represents a speaker’s belief about the frequency of the innovation, based on this one conversation alone. This belief may differ between the two speakers: the values assumed by speakers *i* and *j* we denote as y_i and y_j respectively.

There are clearly many ways in which the two token counts n_i and n_j could be turned into perceived frequencies y_i and y_j. In the original USM of Baxter et al. 2006, 2009, the token counts were simply combined linearly, albeit with different weights. That is, a greater or lesser weight could be given to the interlocutor's token frequency relative to a speaker's own; also the identity of the interlocutor could cause the weight ascribed to his or her utterances to be boosted or suppressed. This latter variation allows for weighted interactor selection.

The linear combination of n_i and n_j, however, precludes replicator selection and nonlinear effects of the type discussed in §§3 and 5, such as frequency-boosting effects. To allow such effects in the more general model introduced here, we instead combine the token counts into a perceived frequency using the formula in 1 for speaker *i* (and the same formula for speaker *j* with all occurrences of *i* and *j* exchanged).

$$(1)\ y_i = H_{ii} f\left(\frac{n_i}{T}\right) + H_{ij} f\left(\frac{n_j}{T}\right)$$

In this expression, H_{ii} is the weight speaker *i* gives to her own utterances, and H_{ij} to those of speaker *j*. Note in particular that we do not require $H_{ij} = H_{ji}$: when these weights are unequal, there is an asymmetry between the speakers *i* and *j* that is characteristic of weighted interactor selection. Finally, in this formula $f(u)$ is some function of u, the fraction of all tokens of the linguistic variable that correspond to the innovation. The specific choice of $f(u)$ depends on the precise mechanism for language change one wishes to model. For example, a linear function recovers the original USM, which is replicator symmetric. Nonlinear forms of $f(u)$ allow us to incorporate the effects of frequency boosting and replicator selection in a systematic way (see §§6.2 and 6.4 below for details).

In order for the language to change, speakers' usage frequencies (the x_i and x_j variables) need to be updated as a consequence of the interaction. In the original USM, the linear rule in 2 was used to update speaker *i*'s frequency.

$$(2)\ x_i(t+1) \propto x_i(t) + \lambda y_i$$

(The same rule with *i* replaced by *j* is used for speaker *j*.) That is, the new usage frequency is the weighted sum of the existing frequency (x_i) and the frequency perceived from the most recent interaction (y_i). If the parameter λ is small, this most recent interaction has a small weight relative to that of earlier interactions, and the language changes slowly over time. Additionally, if λ is large, old tokens are rapidly forgotten, whereas if λ is small they are retained for long periods of time. The relationship between λ and real time is discussed extensively in Baxter et al. 2009. This relationship is not important to us here, however, since we are primarily interested in the shape of the adoption curve, not in how long it takes for an innovation to go to fixation. For clarity of presentation, a constant of proportionality that has no qualitative effect on the resulting dynamics has been omitted from equation 2—the interested reader can find the complete expression in the appendix.

We emphasize that a key property of both the original USM and this extension is that linguistic behavior is VARIABLE, PROBABILISTIC, and DIFFERS BETWEEN INDIVIDUAL SPEAKERS. Each token uttered by a speaker *i* is of the B variant with a probability given by the value of x_i that applies at the time. Additionally, considerable complexity in the social structure is afforded by the sets of interaction frequencies G_{ij} and weights H_{ij}. It is by varying these parameters and the function $f(u)$ that we may systematically explore the consequences of all the change mechanisms identified in §3.

6.1. NEUTRAL EVOLUTION AND NEUTRAL INTERACTOR SELECTION. We begin this study with the case of neutral evolution. Figure 2 shows single trajectories followed by an innovation whose initial frequency across the community is in the range of 10–30%. We remark that all of the details of the innovation process are rolled into the initial condition. That is, we simply state that an innovation has reached some (usually small) frequency by some time $t = 0$, but do not explicitly model the process by which this occurs. This is reasonable for studying the propagation of a new convention that has already been adopted by some minority group within the community (see also §§2, 7.3).

As stated in §5, neutral evolution has the property that the probability of the innovation going to fixation is equal to its initial frequency. Therefore, most innovations, by virtue of their initially low frequency, fail. One such trajectory is shown in Fig. 2. We are, however, primarily interested in how fixation is approached by those rare innovations that do succeed. To generate a sufficiently large number of such trajectories, we used the relatively high initial frequency of 10–30%: two such trajectories are shown in Fig. 2. With lower initial frequencies the shape of the (much smaller number of) propagating trajectories is the same.

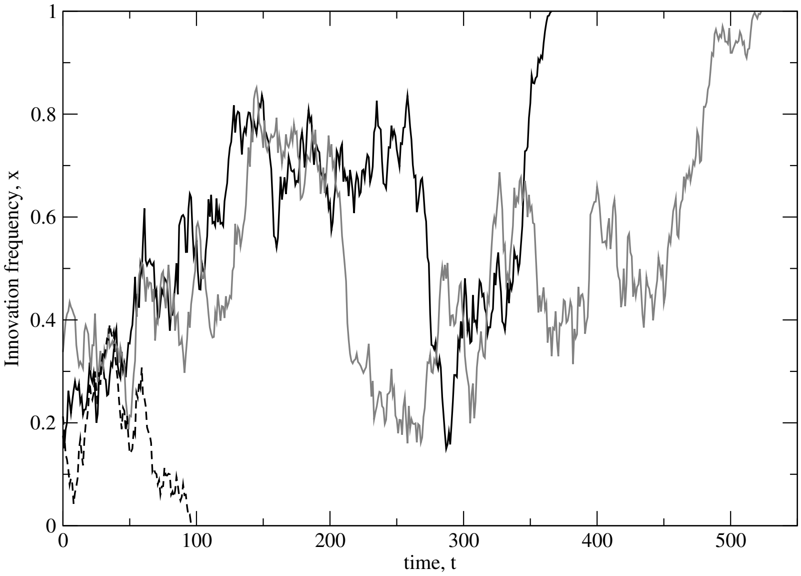

FIGURE 2. Frequencies of an innovation as a function of time in a model with neutral interactor selection and linear transformation rules. The dashed trajectory shows an example of an innovation failing to propagate. The different shading of the two successful trajectories is simply to allow them to be distinguished. These trajectories illustrate the characteristic properties of neutral evolution, namely large fluctuations and a tendency for an upward or downward trend to reverse one or more times before an innovative variant goes extinct or wins out.

These trajectories demonstrate the characteristic property of neutral evolution stated in §5: they fluctuate wildly before fixation is reached. Such fluctuations are not typically evident in the relevant historical data, thereby ruling out neutral evolution as a generic mechanism for the S-curve growth of a new convention. While actual language change trajectories are not perfectly smooth, and represent only approximations to an S-curve (or part of an S-curve), they do not display the sorts of fluctuations found in neutral evolution models. By this we mean that a large number of neutral trajectories we have observed in simulation reach a high frequency before returning to a low frequency for an extended period of time before going to fixation. This sort of fluctuation is not found across the entire set of empirical data supporting the S-curve pattern set out in Table 2.

To obtain Fig. 2, we used a speech community of $N = 80$ speakers, each of whom interacted with eight other speakers, chosen randomly from the whole community at the start of the simulation. This mimics the fact that in reality, any speaker speaks only to a subset of community members. In each interaction, an agent was chosen at random, and then a speaker from the set of eight that he is connected to was chosen. Both agents produced $T = 4$ tokens of a variant in each interaction. Each production event was independent, and the relative frequency with which speaker i produces the innovation (B) was x_i. Likewise, speaker j's was x_j. Then, speaker i formed his perceived frequency via equation 1, and updated his value of x_i using equation 2. Speaker j did the same, but with the labels i and j exchanged in the equations. Replicator neutrality was obtained through the choice $f(u) = u$. In larger speech communities, very similar plots are seen, the main difference being that the time to reach fixation increases linearly with N (see e.g. Baxter et al. 2009 for a discussion of fixation times in a linguistics context).

This is a model of neutral evolution because each speaker has an equivalent role in the community: each speaker interacts with the same number of other speakers, and each interaction occurs with equal frequency. Allowing these interaction frequencies to vary between pairs of speakers yields neutral interactor selection. It has been established (see again Baxter et al. 2009) that with symmetric interaction weights H_{ij} but arbitrary interaction frequencies G_{ij}, the trajectories that result are indistinguishable from the ones shown in Fig. 2.

6.2. NEUTRAL EVOLUTION AND NONLINEAR FUNCTIONS. We use this speech community and set of interaction rules as a baseline against which typical change trajectories arising from different individual behaviors can be compared. As the first example, we examine a nonlinear, but replicator neutral, transformation $f(u)$ that appears in equation 1. Recall that $f(u)$ converts the relative frequency of an innovation in a speaker's utterance to the relative frequency perceived by the speaker. Within this model replicator neutrality applies for any function $f(u)$ if the condition $f(u) + f(1 - u) = 1$ is satisfied (see the appendix). We examine the simplest nonlinear function of this type, given in 3, although the results are similar if we use more complicated functions satisfying the above condition.

$$(3)\ f(u) = u + au(1 - u)(2u - 1)$$

The first term in this formula is the linear rule of the original USM that yields neutral evolution. The second term boosts majority variants and suppresses minority variants if the parameter a is positive: a larger value of a implies a stronger boost. If a is negative, the converse is true: low-frequency variants get boosted at the expense of high-frequency variants. In the biological literature these rules correspond to frequency-dependent selection acting within local subpopulations (see §§3, 5).

By replacing the linear function $f(u)$ with its nonlinear counterpart, equation 3, in the simulations described above (but keeping everything else unchanged), we found the two types of behavior that have been discussed in the literature (see §5). That is, we observed either rapid elimination of the minority variant (when a is positive) or stable coexistence of both variants in equal numbers (when a is negative), as shown in Figure 3. The smoothness of the trajectories and the time required to reach fixation or extinction depend on the size of the community N and the strength of the boost a. In smaller communities, or with smaller boosts, fluctuations become more apparent, and one recovers trajectories like those shown in Fig. 2. This effect is widely observed in related population genetics models (see e.g. Crow & Kimura 1970).

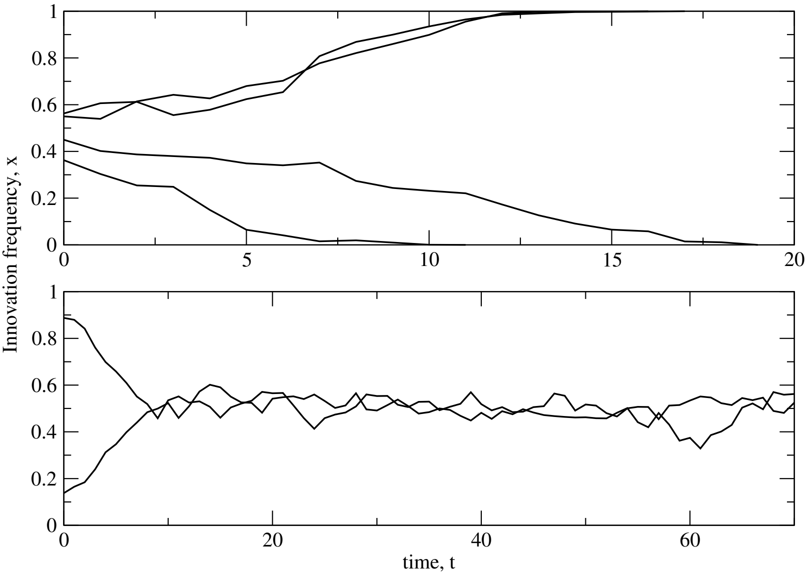

FIGURE 3. Frequencies of an innovation as a function of time in a model with neutral interactor selection and nonlinear sampling rules. In the upper panel, the majority variant was overproduced through a positive value of a in the function $f(u)$ given by equation 3, while in the lower panel, the minority variant was overproduced through a negative value of a in the function $f(u)$. The magnitude of a, whether positive or negative, was 0.02.

6.3. WEIGHTED INTERACTOR SELECTION. We now turn to the case of weighted interactor selection. One way to implement this is by dividing the community into two groups that we call ‘leaders’ and ‘followers’. This we achieve in practice by boosting the weights H_{ij} appearing in equation 1 by a factor α if the speaker is a leader and the listener is a follower (relative to the baseline discussed above). In the original USM (Baxter et al. 2006), which is recovered here through a linear sampling rule $f(u)$, one expects typically that an initially rare innovation should be attracted toward a characteristic frequency, as given in 4, if all n leaders initially use the innovation, and the remaining $N-n$ followers do not (see the appendix).

$$(4)\ x^* = \frac{n\alpha}{n\alpha + (N-n)}$$

This frequency is high if $n\alpha$ is much larger than $N-n$, that is, if the size of the leader group is large, the boost is large, or both. Thereafter, one expects the behavior to revert to that of neutral evolution, that is, undirected fluctuations. As Figure 4 demonstrates, this is what we see: note that the stochastic nature of the trajectories means that the frequency x^* may not be reached on every occasion, or dwelt on for very long. What is always seen, however, is the rapid initial growth of the innovation, the rate of which grows with α. Therefore, if α is large, the innovation is likely to fix: however, the initial slow growth of the innovation characteristic of the S-curve is not seen under these conditions.

When this weighted interactor selection effect is combined with the nonlinear transformation in equation 3, one tends to find that weighted interactor selection is counteracted by the nonlinearity, that is, that trajectories similar to those shown in Fig. 3 are obtained. That is, trajectories typically hit fixation rapidly, but again, not as an S-curve, as the initial growth is very fast and steadily slows down.

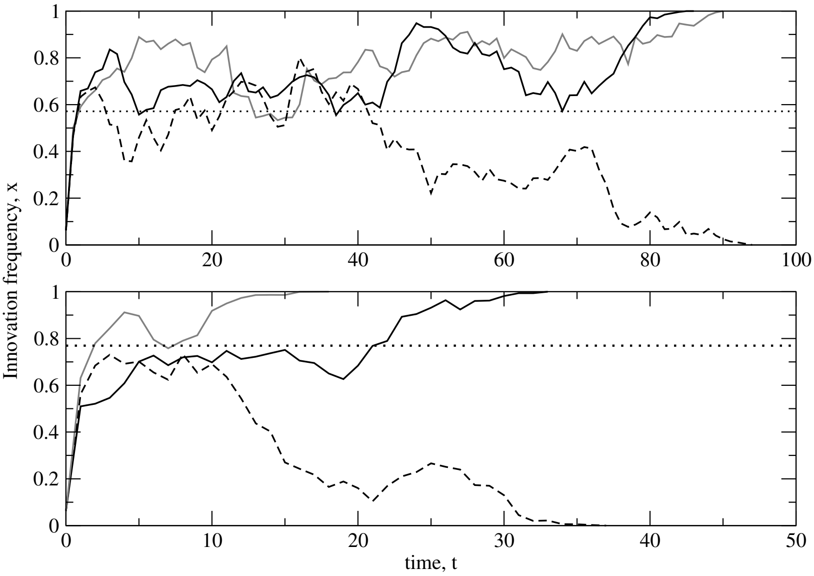

FIGURE 4. Frequencies of an innovation as a function of time in a model with weighted interactor selection implemented by assigning $n = 5$ speakers the status of leaders and with the remaining followers boosting their utterances by a factor $\alpha = 20$ (upper figure) and $\alpha = 50$ (lower figure) more than other utterances. The dotted horizontal line shows the prediction from the original USM for the characteristic frequency that the innovation frequency is initially attracted to before the onset of a neutral evolution regime. As in Fig. 2, we show in each case one case where the innovation fails (dashed trajectory), and two cases where it succeeds (solid trajectories; the different shading is again simply so the curves can be distinguished).

The origin of the initial boost in innovation frequency is the asymmetry between leaders and followers—that is, the weight given by the leaders to the followers differs from the weight ascribed by followers to the leaders. The bigger this asymmetry, the more the effect is enhanced. The extreme case is where speakers only pay attention to the behavior of a single group that comes before them in some sequence: this maps onto Rogers's (1995) framework of adopter categories ('early majority', 'laggards', etc.) where members of each adopter category acquire the innovation from those who are in a more innovative category. Rogers's explanation for the S-curve adoption of an innovation lies in a bell-shaped distribution of adopter categories, with innovators and laggards sitting in the tails of the distribution.

We can implement this setup within the extended USM by dividing speakers up into a sequence of groups, and in each interaction choosing a speaker *i* from the community at random, and then the second speaker *j* randomly from the group that precedes that to which *i* belongs. The interaction weights H_{ij} are set up such that the second speaker never modifies her behavior as a consequence of interacting with the first. (The precise parameter values are given in the appendix: the key point is that whenever speaker *i* does listen to speaker *j*, the value of H_{ij} is as used in the version of the model with neutral interactor selection described above.) A particular feature of this model is that speakers in the first group (innovators) do not listen to anyone else, so never change their behavior. If the innovation is initially confined to this first group of innovators, it is bound to increase over time. One can solve a set of equations that gives the mean innovation frequency, averaged over repeated runs of the dynamics (see appendix). One can also run simulations to examine the trajectory within single runs (i.e. before averaging). We find that the individual trajectories lie close to the mean, and therefore the average trajectory can be identified with a typical trajectory (see Figure 5). If we set up a community with a bell-shaped distribution of adopter categories, we find the innovation does grow smoothly, but that the trajectory is not as S-shaped as one might expect (see Fig. 5, middle panel). The initial period of slow growth is much shorter than is typically observed, and indeed one may question whether the slightly slower initial growth rate predicted by this model would be empirically observable.

The other two panels place the result for the bell-shaped distribution of group sizes into context. The upper panel illustrates what is needed to obtain a more pronounced S-curve, viz, a sequence of adopter categories, each containing progressively more speakers than the previous one. The upper panel has each group twice the size of the previous one—thus just over half of the speakers are laggards, and there is only one innovator. Meanwhile, the lower panel indicates the behavior when all groups are of equal size (barring the innovator group, which contains only a single speaker). Here the initial slow-growth phase is completely absent.

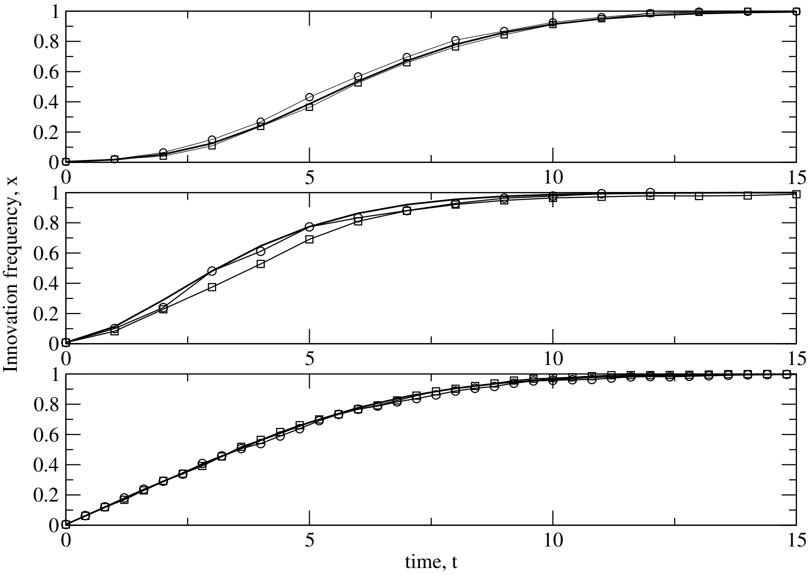

FIGURE 5. Communities subdivided into adopter categories, whereby members of each category monitor only their behavior and that of a speaker from the previous category. In the upper panel, 255 speakers are divided into eight groups, each twice the size of the previous one: 1, 2, 4, 8, 16, 32, 64, 128. In the middle panel, 128 speakers are divided into eight groups with sizes given by the binomial coefficients 1, 7, 21, 35, 35, 21, 7, 1 (this is a natural discrete bell-shaped distribution: it approaches the normal distribution when the number of groups is large). In the lower panel, the innovator group is of size one and the remaining seven groups are of equal size (thirty), making 211 speakers in all. In all panels, the thick line is the average trajectory obtained from a calculation (see appendix), and the points connected by thin lines are from computer simulations.

There is virtually no empirical research on the population structure of adopter groups for language change. Sankoff and Blondeau's (2007) study of the adoption of the uvular variant of (r) in Montreal French by thirty-two speakers interviewed in 1971 and 1984, evenly divided by gender and stratified by age, divides the sample into the following five adopter groups.

- Ten speakers (mostly younger and female) that had already shifted almost entirely to the uvular variant in 1971; Sankoff and Blondeau describe this group as early adopters.
- Seven speakers (mostly younger and male) that shifted to the uvular variant near-categorically by 1984; Sankoff and Blondeau describe this group as later adopters.
- A small group of three speakers that remained with stably variable usage of the apical and uvular variants; this group does not really fit into any of Rogers's adopter categories.
- Two speakers who shifted from near-categorical use of the apical variant to variable use of the uvular variant; this group may correspond to Rogers's late majority.
- Ten speakers (mostly older and male) who continued to use the apical variant near-categorically in 1984; this group corresponds to Rogers's laggards.

Although the number of speakers is small, and possibly not representative, it does not resemble the exponential increase in size of adopter groups from innovators to laggards that is required for an S-curve to emerge solely from weighted interactor selection.

It is possible that the distribution found by Sankoff and Blondeau reflects the fact that they sampled the change in mid-course. Data on the propagation of six grammatical changes documented in the Corpus of Early English Correspondence from Nevalainen et al. 2011 suggests that for the social strata sampled in the corpus, there is no regular social structure of a small number of adopters proceeding exponentially in size through adopter groups to a large group of laggards. Nevalainen and colleagues also examined changes in mid-course, and divided the writers in the corpus into three categories: progressive (leaders), in-between, and conservative (laggards). Nevalainen found that it was not the case that the same writers played the same adopter roles for the six changes in progress: the same individual may be a leader for one change but a laggard for another (Nevalainen et al. 2011:26–32). Also, the numerical distribution of progressive, in-between, and conservative writers for each change differed widely, and none of them resembles an exponential size distribution of adopter groups from progressive to conservative (Nevalainen et al. 2011:24, table 5).

Sankoff and Blondeau's data do exhibit the widely observed sociolinguistic pattern that younger female speakers tend to be early adopters and older male speakers tend to be laggards.[3] Given this correlation of gender and age with receptiveness to innovation, the social structure that is most compatible with an exponential increase in size of adopter groups is one with an inverted age pyramid and a strong gender imbalance favoring males. No known society exhibits either of these traits. This fact suggests that it is highly unlikely that a speech community has the social structure allowing an S-curved trajectory of change with only weighted interactor selection.

Moreover, if the assumption of total asymmetry between adopter categories is reduced, that is, if more innovative speakers are influenced to some degree by less innovative speakers (albeit less so than the converse), one finds the initial slow-growth phase is also washed out to become a longer-lived, almost linear growth (see Figure 6).

Both of these observations indicate that it is highly unlikely that weighted interactor selection gives rise to an S-curve in actual speech communities. We remark that al-

---

[3] Nevalainen and colleagues note that two of their six changes were led by men, but the changes apparently began in administrative and legal language and 'these spheres of life were generally out of reach for women in medieval and early modern times' (Nevalainen et al. 2011:13).

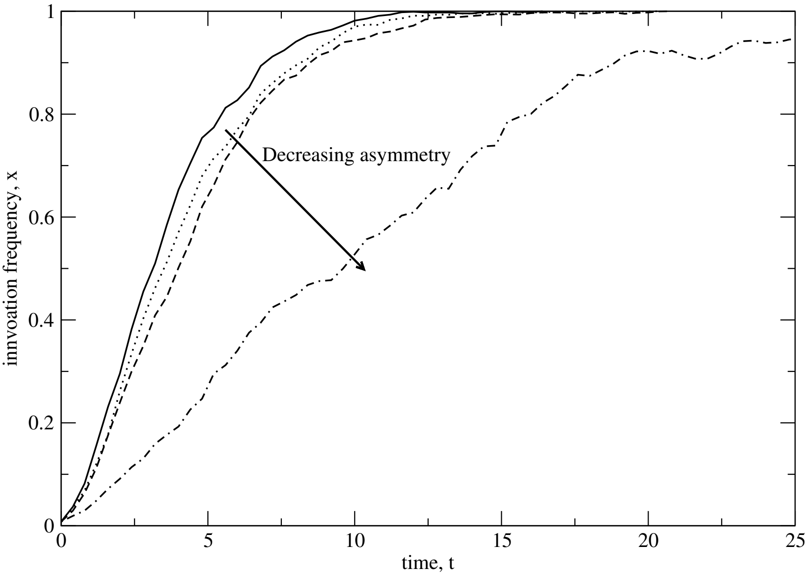

FIGURE 6. The effect of relaxing the assumption of total asymmetry between adopter categories. The bell-shaped distribution of Fig. 5 was used in these simulations. Now a fraction p of interactions, however, have a more innovative speaker listening to a less innovative speaker. That is, $H_{ij}/H_{ji} = q/(1-q)$ where i and j indicate speakers from an adopter category and the immediately more innovative category respectively. $q = 0$ thus corresponds to complete asymmetry, as in Fig. 5. In this figure the values of q used were 0.01, 0.05, 0.1, and 0.3 (from top to bottom). The slightly S-shaped early-growth regime gets washed out as asymmetry decreases.

though we have used fairly small communities and specific network structures for concreteness in this section, the methods of analysis discussed in the appendix are more general and suggest that the findings drawn from our simple models are robust.

6.4. REPLICATOR SELECTION. Taken together with the survey of §5, the results presented above suggest that in order to obtain a convincing S-curve with weighted interactor selection, one needs to tune model parameters quite carefully. By contrast, we have found that an S-shape is easily obtained through replicator selection. Indeed, even the simplest choice of the sampling rule $f(u)$ that implements replicator selection is sufficient to obtain an S-curve. This choice has $f(u)$ linear, but with a slope of greater than one. Then, $f(u) > u$ for all frequencies u between zero and one, and hence the listener perceives the innovation to be at a higher frequency than it was actually produced at, and overproduces accordingly. In Figure 7, we see that the innovation tends to grow even when it starts at low frequency; most noticeable is that the rate of growth is most rapid between a frequency of roughly 20% and 80% (although, since the simulations are stochastic, the S-curve is not as smooth as one might expect: it would become smoother, but faster, as the strength of the replicator selection is increased).

We remark that this linear boosting of the innovation frequency is not the standard way to implement classical fitness in population genetics models, which normally involve a nonlinear function that leads to a logistic growth of the innovation (see e.g. Crow & Kimura 1970). A prediction for the form of this curve can be obtained by invoking some approximations about the distribution of usage frequencies (see appendix). These predictions are shown as thin lines in Fig. 7 and agree rather well with those obtained from the simulation. The form of these curves is not, in general, logistic: the exact shape depends on the number of tokens produced by a speaker in each interaction. The initial rate of growth is, however, independent of the number of tokens that are produced, thus retaining the essential component of the S-curve trajectory. This suggests that, from the point of view of obtaining an S-curve, the precise nature of the preference for the innovation is of less importance than the fact that some form of replicator selection is active.

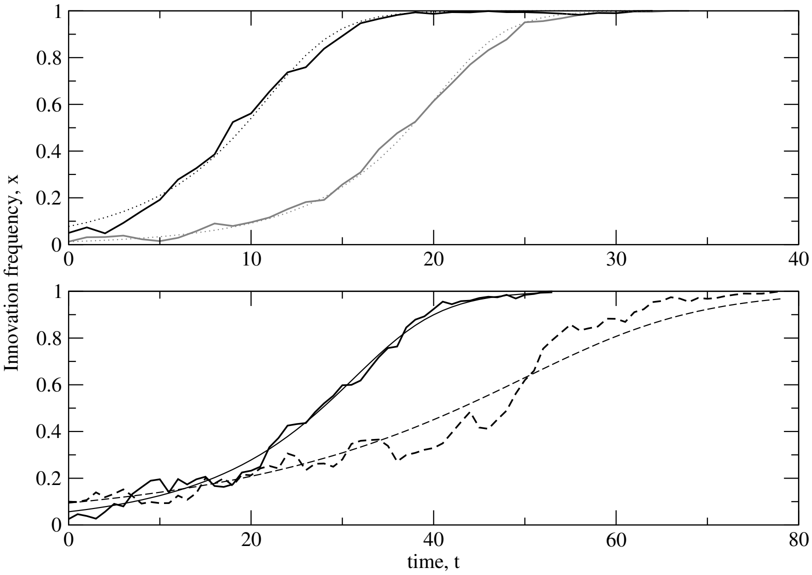

FIGURE 7. Model with replicator selection implemented by boosting the innovation by an amount proportional to its frequency, that is, through the rule $f(u) = (1 + b)u$. In the upper panel, $b = 0.001$ for all speakers. The two thick (and slightly noisy) curves come from two different stochastic simulations with slightly different initial frequencies of the innovation; the smooth thin curves are predictions from a deterministic theory described in the appendix. The only variable parameter in this theory is the initial frequency of the innovation: the light shaded curve has a lower initial frequency than the heavy shaded curve. In the lower panel, some fraction of the speakers are positively disposed to the innovation and the remainder are negatively disposed: for both types of speakers the slope of $f(u)$ is the same ($b = 0.001$), but the intercept is different (see appendix). The fraction of positively disposed speakers was 70% for the solid curves, and 60% for the dashed curves. Again, the thick curves are from stochastic simulations and the thin curves from theory (see appendix).

The most interesting question one can ask about replicator selection is how uniformly it needs to be applied across the community for sustained directed growth of the innovation to result. We examined this numerically by setting the slope of the linear function $f(u)$ randomly on a per-speaker basis so that some were slightly greater than one, and some slightly less than one. With the strength of replicator selection we used, we found that an S-curve was still reliably obtained if 70% of the community were pro-innovation, and the remaining 30% were anti; once this dropped below about 60%, the innovation tended to die out rapidly with a high probability, except if the initial usage frequency happened to be sufficiently large, as shown in Fig. 7. This demonstrates that while a majority of the community needs to boost the innovation, it is not necessary for all speakers to be biased toward it—indeed, some may be biased toward the convention without preventing fixation of the innovation along an S-curve trajectory.

7. DISCUSSION AND CONCLUSIONS.

7.1. THE MODEL AND THE THEORIES. The results from the modeling imply that replicator selection, that is, differential weighting of linguistic variants, is almost certainly an essential mechanism for language changes that follow an S-curve, specifically, when a competing variant replaces a previously established conventional variant. Since the language changes of this type that have been quantitatively documented over their lifetimes follow an S-curve or are compatible with an S-curve, our results suggest that most if not all such changes involve differential weighting of linguistic variants. That is, the Sturtevant-inspired mechanism plays a central role in language change.

As noted at the end of §6, it is not necessary for all speakers to give the same differential weighting to the ultimately successful linguistic variant. Furthermore, these weightings do not need to remain constant over time. Indeed, Vulanović and Baayen (2007) have argued that Ellegård's (1953) data for the rise of English periphrastic *do* are best fit by a weighting that takes one value at the start of the change, and jumps to a different value at some point during the change. The key point is that variation of the weightings between speakers and over different time periods seems less important in terms of obtaining an S-curve than the fact that variants are differentially weighted.

It is possible to simulate an S-curved trajectory of language change without replicator selection. However, it requires averaging over multiple trajectories (as was done by Reali and Griffiths (2010)) or a very special condition of weighted interactor selection: each successive following group (early adopters, late adopters, laggards, etc.) is exponentially larger in population, and weights only the language use of the immediately 'preceding' adopter group in the sequence of diffusion. It is risky (to say the least) to assume that a widespread empirical pattern like the S-curve in language change is due to a very specific combination of values for the variables in a model, such as the exponentially increasing size of adopter groups. And the correlation of adoption of an innovation with gender and age, combined with the typical age and gender structure of human societies, suggests that this specific combination of values does not hold for actual speech communities.

What the two mechanisms we have found that generate an S-curve have in common is that they exploit the replicator or interactor asymmetry to drive a directed change through the community. In mathematical terms, this asymmetry can ensure that the existing convention is an unstable equilibrium and the innovation is a stable equilibrium. An S-curve change is then typically to be expected (Strogatz 1994).

In the simulations we have presented, we have used mostly small speech communities with relatively simple social structures for illustrative purposes. We believe, however, that our results are robust to both increased community size and complexity. In models with no selection or only neutral interactor selection, increased community size is known to slow down the rate of fixation, and network structure has no effect on the dynamics (see e.g. Baxter et al. 2009). In models with weighted interactor selection alone, all systematic effects originate in the network structure. For these to be pronounced, however, one typically needs the somewhat stark subdivision into adopter groups, as was the case in §6.3. If, for example, one introduces further structure that leads to increased contact between individual members of the different groups, the innovation diffuses more easily between groups and weakens the social network effects. When replicator selection is present, its effect tends to dominate that of network structure, a fact whose implications we discuss in more detail below.

In an effort to go beyond observations drawn from specific simulation conditions, we have attempted to develop more general mathematical arguments to understand the robustness of our results. These arguments are presented in detail in the mathematical appendix. These are incomplete and inexact, in that they rely on certain simplifying assumptions that allow the general consequences of interactor and replicator selection, as defined in §3, to be analyzed. We have found that certain predictions arising from this mathematical approach, however, agree very well with simulation data despite the simplifying assumptions: such agreement can be seen in Figs. 4 and 6. We are thus confident that the general patterns we have reported here are not restricted to the very specific cases discussed in §6.

A second striking result from our modeling work is that the Bloomfield-inspired mechanism, neutral interactor selection—that is, the frequency of interaction of individuals and tie strength—does not affect the results. Neutral interactor selection by itself produces only random fluctuations; that is, it does not produce an S-curved trajectory of language change. Moreover, replicator selection leads to an S-curved trajectory no matter what pattern of neutral interactor selection—that is, no matter what social network structure—is assumed.

These results suggest that any theory of language change must put some kind of differential weighting of linguistic variants in a central role. Trudgill's theory of new-dialect formation eschews any appeal to differential social weighting of linguistic variants, and places all of the explanatory burden on social network structure. If the variants that become the conventions of the new dialect arrive at that position via an S-shaped trajectory, then it is likely that Trudgill's theory of new-dialect formation will have to include differential weighting of linguistic variants.

The Milroys' theory of language change assumes that differential weighting of variants is part of the explanation. They argue, however, that it plays a minor role in language change; the more important role is played by network structure. Our results suggest that the importance of the two factors, at least for the overall trajectory of language change, is the other way around: differential weighting is the essential factor.

Labov's theory also assumes that differential weighting of variants is part of the explanation. Although he invokes social network structure as a possible factor, as noted in §§1 and 3, it does not play a central role in the model outlined at the end of his 2001 book. Hence Labov's theory, as well as Mufwene's, is the one most compatible with the results described here.

The models we used tell us only that a differential weighting of variants drives the overall propagation process. It does not tell us what that weighting corresponds to in the linguistic and social world, or where the weighting comes from. It is generally assumed in sociolinguistic research that the differential valuation of linguistic variants is social: prestige, covert prestige, acts of social identity, and so forth. We agree that some such factor or combination of factors plays a role, but only empirical studies of social values and language use can determine which factor(s) determine the weighting.

7.2. REPLICATOR SELECTION AND THE OTHER MECHANISMS. We emphasize that our results do not exclude the possibility of a role for social network structure (neutral interactor selection) and distinct roles of social groups in language change (leaders and different groups of followers/adopters; weighted interactor selection). At the end of §4, we observed that the overall structure of the trajectory of a language change was an S-curve, no matter how it was propagated through grammatical contexts, words, speakers, texts, geographical regions, or social classes. This overall trajectory appears to be determined by differential weighting of variants (replicator selection). The microstructure of the process—how it is diffused through subparts of linguistic and social context—may be driven by social network structure and weighting of the language use of different subgroups of the population.

When we combine replicator and weighted interactor selection, we find in the leaders-follower model that the initial growth rate of the innovation increases with the weight assigned to the leaders' utterances, so becomes less S-shaped the more strongly the interactor is weighted. Since in the Rogers-type model the asymmetric characteristic of weighted interactor selection is maximal, the addition of replicator selection does not change the trajectories shown in Fig. 5. Likewise, when one combines replicator selection with the nonlinearity of equation 3, the stronger effect tends to dominate. For example, if the majority variant is overproduced, and the innovation is overproduced, then the innovation can be rapidly eliminated if the bias against the minority variant outweighs the preference for the innovation. Likewise, if the majority variant is underproduced, one can find a coexistence of both variants but in unequal numbers.

We intend to explore the role of neutral interactor selection and weighted interactor selection in relation to replicator selection in future modeling. To some extent, however, the microprocesses proceed independently of the overall S-shaped trajectory of propagation, which is determined by the differential weighting of variants.

7.3. WHERE DOES THE DIFFERENTIAL WEIGHTING OF VARIANTS COME FROM? A plausible hypothesis, proposed independently by Croft (2000:186) and Labov (2001:517–18), is that the differential weighting of a variant originates in the differential weighting of speakers favoring that variant in their language use. For whatever reason, a speaker or group of speakers weights the language use of other (groups of) speakers differentially. That speaker/group then observes that a particular variant is probabilistically associated with the relevant speaker/group, and transfers the weighting from the speaker/group to the variant. For example, Labov writes: 'Outliers [variants] … are heard as characteristic of younger speakers and emphatic, less monitored speech, deviant from the accepted norm of older speakers' (2001:517). Once that step is made, replicator selection is operating, as the outlier variants are now associated with a social value. At that point, it is the variant that possesses the (social) value, no matter who produces it. That is, the weighting assigned to the interactor (the speaker) is transferred to the replicator (the linguistic variant). This is a plausible hypothesis on the assumption that the valuation of the variant is based on the same social factors that speakers use to value the language use of other speakers.

The Labov/Croft hypothesis represents a different relationship between the generation of variation and its propagation than is found in standard accounts of biological evolution. In biological evolution, variants generated by mutation or sexual recombination have an inherent fitness value. Differences in fitness values of variants immediately lead to their propagation/extinction via replicator selection. In language change, according to the Labov/Croft hypothesis, variants are generated by processes that are ultimately phonetic in origin (for sound change) or semantic/pragmatic (for grammatical change). The variants do not have inherent social values. These values must be acquired by some process, such as the one Labov and Croft propose, before replicator selection can operate on the variants.

The Labov/Croft hypothesis may be informed by the following observations about our simulations. We noted in §7.2 that the relative strength of weighted interactor selection and replicator selection affects the outcome: if weighted interactor selection is much stronger than replicator selection, then an S-curve does not occur. The relevant quantity is the mean rate of increase of a variant divided by its overall frequency. With replicator selection defined as it is here, this ratio is the number b that is referred to in the caption of Fig. 7 and characterizes the strength of replicator selection. (It also appears in equation A21 of the mathematical appendix.) For weighted interactor selection, we have to take the initial condition into account. If variants are randomly distributed among speakers, weighted interactor selection does nothing: it only leads to propagation when there are separate conventions within groups that are weighted differently.

For a simple leaders-followers model of the type proposed by sociolinguists (§3.3), with leaders using the innovation and followers the existing convention, the ratio defined above is $(\alpha - 1)h$ where α is the popularity of the leaders. The variable h was set to 1 in our simulations, so the numbers to compare are $\alpha - 1$ and b. Our simulations suggest that the effect of replicator selection is more visible than that of weighted interactor selection when $b > \alpha - 1$. The simulations with weighted interactor selection use $\alpha = 20$. When α is set this high, the weighting for replicator selection must be set so high that fixation occurs very rapidly, along a curve that has no obvious shape (because it is too short in time). By contrast, reducing α means that weighting of interactors is decreased; the end point of this process is equal weighting of interactors, that is, neutral interactor selection. At this end point, we recover the behavior described in §6.4, that is, an S-shaped trajectory (with minor fluctuations) enforced by replicator selection. Simulation results (not shown) suggest that when weighted interactor selection and replicator selection are of similar strength, weighted interactor selection works to prevent the innovation from going extinct to begin with, and replicator selection takes over to bring fixation along an S-shaped trajectory.

The key point is that the additional weighting given to a variant through weighted interactor selection is directly comparable to the additional weighting given by replicator selection. Hence the hypothesis proposed by Croft and Labov is a plausible scenario in the model: change starts through weighted interactor selection, speakers transfer the interactor's weighting to the replicator, and change proceeds to fixation through replicator selection (assuming no further changes in replicator weighting).

APPENDIX

A1. MATHEMATICAL DEFINITIONS OF SELECTION MECHANISMS. As stated in §3 of the main text, the different selection mechanisms we have identified—neutral interactor selection, weighted interactor selection, and replicator selection—are naturally formulated in terms of symmetries in a model. Here, we introduce mathematical definitions of these selection mechanisms that can be applied to any model, not just the utterance selection model of Baxter et al. 2006, 2009.

A1.1. REPLICATOR SELECTION. We use the term REPLICATOR SELECTION to imply an asymmetry in the dynamics under exchange of different variants. By this we mean that if there is any aspect of the dynamics that depends on the labeling of the variants, replicator selection is operating. More precisely, consider the case of two variants A and B. Let us also denote the initial distribution of these variants (i.e. assignment of usage frequencies to individual speakers) as X. After some time t a distribution X' is reached with probability $P(X',t|X)$. Now consider the initial condition $\bar{X}$ with all A and B variants exchanged. Then, the similarly exchanged distribution $\bar{X}'$ is reached with probability $P(\bar{X}',t|\bar{X})$. We say replicator selection is acting if A1 holds for any X, X', and t.

$$\text{(A1)} \quad P(\bar{X}',t|\bar{X}) \neq P(X',t|X)$$

This definition can be extended to more variants by considering all possible exchanges of variant labels.

A1.2. NEUTRAL AND WEIGHTED INTERACTOR SELECTION. Interactor selection relates to a symmetry under exchange of speaker identities, rather than variant identities. The qualitative idea we want to formalize is speaker i having the 'same influence' on speaker j as speaker j has on speaker i. One way to express this is through the usage frequency x that a given speaker is using a particular variant. Suppose that initially, only speaker j is using this variant (we call this the innovation). We can then ask for the probability that speaker i is using the innovation with frequency x after time t given that it was innovated by speaker j: let us denote this as $P_i(x,t|j)$. We can also consider the complementary case, that is, where speaker i is the innovator, and write down the probability $P_j(x,t|i)$ that speaker j is using the innovation with frequency x after time t. We say that a model exhibits NEUTRAL INTERACTOR SELECTION if A2 holds for all x, t, and i.

(A2) $P_i(x,t|j) = P_j(x,t|i)$

Otherwise the model exhibits weighted interactor selection.

Neutral evolution is a special case of neutral interactor selection where all speakers are completely equivalent—that is, there is no difference at all when speakers are exchanged. Neutral interactor selection admits the possibility that different speakers interact with each other with different frequencies. In Baxter et al. 2009 it was found that certain results were independent of network structure when neutral interactor selection was operating.

A2. GENERAL CONSEQUENCES OF THE SELECTION MECHANISMS. In principle one would like to use the definitions of replicator selection (A1) and interactor selection (A2) to determine constraints on the shape of a change trajectory. There are many steps, however, that lie between these definitions and an expression for the overall usage frequency of an innovation as a function of time. For example, to work out the probability $P_i(x,t|j)$ that speaker i is using a variant introduced by speaker j, we would in principle need to know all possible routes between speakers i and j that the innovation may have followed, how the linguistic behavior of all agents along these paths is characterized and affected by their interlocutors, and so on. Nevertheless, by making some simplifying assumptions we can take some steps toward this goal. In particular, we can argue that—subject to these assumptions—an S-curve trajectory depends on either replicator selection or possibly weighted interactor selection to be present, as seen in the simulations in the main text.

A2.1. REPLICATOR SELECTION. Let us suppose that the dynamics are such that after some time t, the rate of change of the innovation frequency in the entire community, x, is a deterministic function of its frequency, as in A3.

(A3) $\frac{\mathrm{d}}{\mathrm{d}t}x = F(x)$

While the rate of change of x will often depend on HOW the innovation is distributed among its speakers, we have encountered in our simulations certain combinations of network structures and interaction rules where this equation seems to be a reasonable approximation. The key point is that the symmetry implied by an absence of replicator selection places strong constraints on the form of $F(x)$.

Consider the frequency of the convention $\bar{x} = 1 - x$. Clearly, whenever x increases, $\bar{x}$ decreases by the corresponding amount, and we have A4.

(A4) $\frac{\mathrm{d}}{\mathrm{d}t}\bar{x} = -\frac{\mathrm{d}}{\mathrm{d}t}x = -F(x)$

If we denote the rate of increase of the convention as $\bar{F}(\bar{x}) = \bar{F}(1-x)$, we have A5.

(A5) $\bar{F}(1-x) = -F(x)$

In the absence of replicator symmetry, the rate of increase of the convention or innovation as a function of its frequency must be given by the same function, since we have stated that the dynamics must be invariant under relabeling of variants. That is, $\bar{F}(u) = F(u)$ and hence we have A6 if replicator selection is not operating.

(A6) $F(1-x) = -F(x)$

A consequence of this relation is that when the innovation frequency is 50%, we require that $F(\frac{1}{2}) = -F(\frac{1}{2})$. The only possible solution of this equation is that $F(\frac{1}{2}) = 0$. Thus coexistence of the innovation and convention in equal numbers is an equilibrium of the dynamics. The only remaining question is whether the equilibrium is stable or unstable. Stable coexistence is obtained if the minority variant is overproduced, and the majority variant underproduced; by contrast, if the majority variant is overproduced, the minority variant will be eliminated. This accounts for two of the patterns frequently seen in models, as discussed in §§5 and 6 of the main text. The S-curve requires that an innovation be boosted at low and high frequencies: equation A6 indicates that this is impossible within the approximation described here. Therefore one would expect to have to break the symmetry between variants to obtain sustained growth of an innovation over an extended period.

A2.2. INTERACTOR SELECTION. We now examine the effects of interactor selection in the absence of replicator selection. Let us first assume that the social network is such that an innovation can spread from its initial location very rapidly. Typically this phenomenon is observed on networks with the small-world property (Watts & Strogatz 1998), which is where the number of steps between any two agents on the network is considerably smaller than the number of agents on the network. Popularly, this is referred to as 'six degrees of separation' and is often assumed to be a property of social networks. One possibility then is that the probability of any given speaker j using an innovation introduced by speaker i depends on the identity of the original innovator, but not of the speaker later using it. The picture here is that initially, an innovation may spread or go extinct; but if it does spread, it spreads across the community so that everyone uses it with a similar probability. The identity of the innovator may affect the initial probability of spreading, but not the subsequent dynamics.

If neutral interactor selection is operating within this scenario, it turns out that the identity of the original innovator has no effect on the probability that it is used with a frequency x by any chosen member of the community. This is because, if i has innovated, we have assumed that $P_j(x,t|i)$ defined above is independent of j: $P_j(x,t|i) = f_i(x,t)$. From the mathematical definition of neutral interactor selection (A2), however, we also have that $P_i(x,t|j) = f_j(x,t) = f_i(x,t) = f(x,t)$, for any i and j. That is, the probability that any speaker is using the innovation has the same form, no matter the identity of the original innovator.

There is one case that is widely observed where the assumption we have made here—that the usage frequency of an innovation is the same for every speaker—is valid. That is when the innovation has gone extinct or to fixation. Then, by definition, $P_j(x,t|i) = 0$ or 1 for all speakers j. Neutral interactor selection then implies that the probability that the innovation replaces the existing convention under neutral interactor selection does not depend on the identity of the innovator.

In the absence of replicator selection, the function $f(x,t)$ that gives the probability distribution of the usage frequency x for every speaker under neutral interactor selection is constrained by the replicator symmetry described in the previous subsection. Then, one does not expect the sustained growth of an innovation. However, under weighted interactor selection, where the probability that a given speaker uses the innovation CAN depend on the identity of the innovator, there is the possibility for the innovation to grow in frequency due to the social dominance of the speakers using the original innovation. This is observed, for example, in the simple models of weighted interactor selection discussed in the main text.

Together, the general—but approximate—arguments of this section are suggestive that a necessary criterion for sustained growth of an innovation is replicator selection or certain types of weighted interactor selection. In turn, sustained growth of an innovation is a necessary criterion for an S-curve—but may not be sufficient. A more precise understanding of the conditions for both sustained growth or an S-curve would be worth pursuing.

A3. DEFINITION OF THE EXTENDED UTTERANCE SELECTION MODEL. We use an extension to the utterance selection model introduced in Baxter et al. 2006 to allow all of the above selection mechanisms to be implemented in a systematic way. The model is defined as follows.

- The speech community comprises N speakers, labeled $i = 1, 2, \dots, N$. Associated with each speaker is a frequency x_i, a number between zero and one, that specifies the probability that speaker i uses the innovative variant of a linguistic variable. The conventional variant is used with probability $1 - x_i$.
- The system evolves through a sequence of repeated interactions. In each interaction, a pair of speakers is chosen to interact. The probability that speakers i and j interact is given by G_{ij}, normalized so that $\sum_{i<j} G_{ij} = 1$.
- Both speakers i and j produce T tokens of the linguistic variable. Each is a token of the innovative variant with probability x_i and x_j respectively. We use the numbers n_i and n_j to denote how many tokens of the innovation have been produced in the interaction by each speaker.
- In their role as listeners, speakers i and j convert the uttered token numbers n_i and n_j into a perceived usage frequency of the innovation through the formula in A7, where H_{ij} is the weight given by speaker i to utterances produced by speaker j, and $f(u)$ is a function that transforms the actual token frequency into a perceived token frequency.

  $$\text{(A7)}\quad y_i = (1 - H_{ij})f\left(\frac{n_i}{T}\right) + H_{ij}f\left(\frac{n_i}{T}\right)$$

  The quantity y_j is obtained through the expression obtained by exchanging i and j.

  In the original USM, $f(u) = u$; various selection mechanisms and nonlinearities can be implemented in this model through different choices for $f(u)$ as described in the main text and below. In the original USM, the factor $(1 - H_{ij})$ was not present; here it is included for mathematical convenience in the following. Since H_{ij} is normally small, this is not an important difference.
- Speaker i updates her usage frequency x_i through formula A8.

  $$\text{(A8)}\quad x_i' = \frac{x_i + \lambda y_i}{1 + \lambda}$$

  Speaker j does likewise (with the replacement $i \rightarrow j$ everywhere in this expression). The parameter λ is intended to be small, since the usage frequency is expected to change only slightly as the result of a single interaction (see Baxter et al. 2006, 2009).

  The specific form of the denominator in the previous equation ensures that the probabilities of using the convention and the innovation sum to one. To see this, we introduce the frequency of the convention as used by speaker i, $\bar{x}_i$, and the function $\bar{f}(\bar{u})$ that transforms the actual convention frequency to its perceived frequency. If we have $x_i + \bar{x}_i = 1$ before the interaction, then we have A9, where $u_i = \frac{n_i}{T}$, $u_j = \frac{n_j}{T}$, and $\bar{u} = 1 - u$.

$$\text{(A9)}\quad x_i' + \bar{x}_i' = \frac{1}{1+\lambda}\left(x_i + \bar{x}_i + \lambda[1 - H_{ij}][f(u_i) + \bar{f}(\bar{u}_i)] + \lambda H_{ij}[f(u_j) + \bar{f}(\bar{u}_j)]\right)$$

This expression equals one if A10 is true.

$$\text{(A10)}\quad \bar{f}(\bar{u}) = 1 - \bar{f}(u)$$

We effectively use this criterion to DEFINE the transformation of convention frequencies once the function $f(u)$ has been specified. One assumption that is made in the main text is that neither variant is ever spontaneously generated: therefore $f(0) = \bar{f}(0) = 0$. Consequently, $f(1) = \bar{f}(1) = 1$. For intermediate frequencies, the form of $f(u)$ is unconstrained, other than it must produce a perceived frequency that lies between zero and one.

The natural unit of time in the model is N/λ^2 interactions. This is the unit of time used in each of the figures in the main text.

A4. SIMULATION CONDITIONS. We summarize the parameters used in the simulations of the USM to obtain the trajectories plotted in §6 of the main text. We also sketch the derivations of equation A6 and the analytical curves plotted in Fig. 4.

Neutral interactor selection and no replicator selection. The data for Fig. 1 were obtained by constructing a network in which each of the N speakers was connected to a random subset of k other speakers in the community. These connections were formed by initially ascribing a set of k 'sockets' with each speaker. Then, the network is built up by repeatedly connecting together a pair of unconnected sockets at random, subject to the constraint that the sockets must belong to distinct speakers who are not already connected. Then, for any pair of connected speakers i and j, $G_{ij} = 2/(NK)$. Neutral interactor selection is achieved by setting all H_{ij} equal and taking $f(u) = u$. Following Baxter et al. 2006, 2009, we set $H_{ij} = \lambda h$. We used an initial condition in which each speaker was designated a categorical user of the innovation ($x_i = 1$) with probability p, and of the convention ($x_i = 0$) otherwise.

The trajectories shown in Fig. 1 were obtained using the following parameter choices: $N = 80$, $k = 8$, $p = 0.2$, $\lambda = 0.01$, $h = 1.0$, and $T = 4$.

The trajectories shown in Fig. 2 were obtained using the same set of parameters and network structure, but after replacing the linear function $f(u)$ with the nonlinear function in A11.

$$\text{(A11)}\quad f(u) = u + au(1 - u)(2u - 1)$$

This is replicator neutral because $\bar{f}(u) = 1 - f(1 - u)$ has the same functional form. We used $a = 0.02$ to model the case where the majority variant is systematically overproduced, and $a = -0.02$ to model the case where the majority variant is systematically underproduced.

Weighted interactor selection: Followers and leaders. The same network structure as described above was used to model a community subdivided into a group of leaders and a group of followers. The first n of the N speakers were designated leaders, and the remainder followers. The utterances of leaders were boosted by followers. Thus the H_{ij} parameters take the two possible values in A12.

$$\text{(A12)}\quad H_{ij} = \begin{cases} \lambda h\alpha & \text{if } i > n \text{ and } j \leq n \\ \lambda h & \text{otherwise} \end{cases}$$

Here α is a parameter that quantifies the strength of the weighted interactor selection.

If λ is small, this model falls into the class described by Baxter and colleagues (2008), who identified the characteristic frequency x^* discussed in the main text (denoted ξ in Baxter et al. 2008) as in A13.

$$\text{(A13)}\quad x^* = \sum_{i=1}^{N} Q_i x_i(0)$$

Here $x_i(0)$ is the initial usage frequency of the innovation for speaker i, and Q_i is the solution of the set of equations in A14, subject to the normalization $\sum_i Q_i = 1$.

$$\text{(A14)}\quad \sum_j (Q_i h_{ij} - Q_j h_{ji}) = 0$$

For this model, the solution of these equations is A15.

$$\text{(A15)}\quad Q_i = \begin{cases} \dfrac{\alpha}{n\alpha + N - n} & \text{if } i \leq n \\ \dfrac{1}{n\alpha + N - n} & \text{if } i > n \end{cases}$$

In the simulations, we took all leaders initially to be categorical users of the innovation ($x_i(0) = 1$ for $i \leq n$) and followers to be categorical users of the convention ($x_i(0) = 0$ for $i > n$). Performing the sum A13 yields equation 6 in the main text. All parameters were the same as stated above for the neutral interactor selection simulations. We took $\alpha = 20$ and $\alpha = 50$ to demonstrate the effect of the weighted interactor selection.

Weighted interactor selection: Rogers-type models. In the main text we discuss a model invoking weighted interactor selection inspired by a picture of innovation diffusion due to Rogers (1995). This has a different network structure from that described above. Speakers are divided into groups, $g = 1, 2, \ldots, m$, with n_g speakers in group g. Interactions take place between members of neighboring groups, but speakers from group g listen only to the utterances of speakers from group $g - 1$. The model was set up such that each speaker from groups 2, 3, … , m had an equal probability of listening to the speaker from the preceding group in each interaction. In practice this is how it was simulated. It can be formally represented, however, using the quantities G_{ij} and H_{ij} via A16 and A17, in which g_i and g_j are the groups to which speakers i and j belong.

$$\text{(A16)}\quad G_{ij} = \begin{cases} \dfrac{1}{N - n_1}\,\dfrac{1}{n_{g_i}} & \text{if } g_i = g_j - 1 \\ \dfrac{1}{N - n_1}\,\dfrac{1}{n_{g_j}} & \text{if } g_i = g_j + 1 \end{cases}$$

$$\text{(A17)}\quad H_{ij} = \begin{cases} \lambda h & \text{if } g_i = g_j + 1 \\ 0 & \text{otherwise} \end{cases}$$

In the simulations, we used an initial condition in which all n_1 members of group 1 used the innovation categorically (these are the innovators). The speakers from the other groups initially all used the convention categorically. As in the other simulations, $\lambda = 0.01$, $h = 1.0$, and $T = 4$. We used two different group-size distributions. To mimic the bell-shaped distribution suggested by Rogers (1995), we took the group sizes to be binomial coefficients (since this approximates a normal distribution when the number of groups is large). For the case of seven groups discussed in the main text, this yields $n_g = 1, 7, 21, 35, 35, 21, 7, 1$. We contrasted this with the case where the group sizes doubled as the distance from the innovator increased: $n_g = 1, 2, 4, 8, 16, 32, 64, 128$.

In Fig. 4, the solid curve shows the solution of the equations governing the rate of change of the mean frequency of the innovation within each group. This solution is a little involved: we sketch the main steps here for the mathematically expert reader. By performing an appropriate average over the Fokker-Planck equation presented in Baxter et al. 2006, one can show that the mean frequency of the innovation $\tilde{x}_g$ within group g is obtained from the solution of the differential equation A18.

$$\text{(A18)}\quad \frac{\mathrm{d}}{\mathrm{d}t}\,\tilde{x}_g(t) = \frac{h}{1 - n_1/N}\,[\tilde{x}_{g-1}(t) - \tilde{x}_g(t)]$$

This is a set of coupled differential equations: to find an explicit expression for $\tilde{x}_g$ one needs to know the form of $\tilde{x}_{g-1}$. These coupled equations can be solved by using Laplace transforms, which is a standard technique for solving such equations (see e.g. Boas 1983). In our simulations, the combination $h/(1 - n_1/N)$ was close to one; the function $\tilde{x}_g$ is then approximately given by the inverse Laplace transform of the function A19 for groups $g \geq 2$.

$$\text{(A19)}\quad \frac{1}{s}\,\frac{1}{(1+s)^{g-2}}$$

This inversion can be computed using a computer algebra package (e.g. Mathematica or Maple). The mean usage frequency of the innovation over all members of the community is then obtained using the weighted sum A20.

$$\text{(A20)}\quad \tilde{x}(t) = \sum_g \frac{n_g}{N}\,\tilde{x}_g(t)$$

It is this expression that is plotted as the solid curves in Fig. 4.

Replicator selection. Replicator selection demands that the innovation is systematically overproduced for all frequencies. The simplest way to achieve this is with a linear function $f(u)$ whose slope exceeds one. Since for high frequencies, this would cause $f(u)$ to exceed one, we set $f(u) = 1$ in this high-frequency range, where the parameter $a > 0$.

$$\text{(A21)}\quad f(u) = \begin{cases} (1 + b)u & \text{for } 0 \leq u \leq \dfrac{1}{1 + b} \\ 1 & \text{for } \dfrac{1}{1 + b} \leq u \leq 1 \end{cases}$$

This implies that the perceived frequency $\bar{f}(u)$ of the convention is reduced to zero if the actual frequency u falls within the low-frequency range.

$$\text{(A22)}\quad \bar{f}(u) = \begin{cases} 0 & \text{for } 0 \leq u \leq 1 - \dfrac{1}{1 + b} \\ (1 + b)u & \text{for } 1 - \dfrac{1}{1 + b} \leq u \leq 1 \end{cases}$$

Across the whole range of u, the convention is underproduced relative to its actual frequency. In the main text we make reference to a case where some speakers are negatively disposed toward the innovation: in this case we use this second form of the frequency transformation for the innovation (rather than the convention).

In the simulations, we used the same network structure and parameters as for the case of neutral interactor selection, albeit with a reduced initial frequency of the innovation (we used values of p between 0.01 and 0.1). The value of b used in the simulations to obtain Fig. 5 was 0.001.

REFERENCES

Abrams, Daniel M., and Steven H. Strogatz. 2003. Modelling the dynamics of language death. *Nature* 424.900.

Altmann, Gabriel; Haro von Buttlar; Walter Rott; and Udo Strauss. 1983. A law of change in language. *Historical linguistics*, ed. by Barron Brainerd, 104–15. Bochum: Brockmeyer.

Ayala, Francisco J., and Cathryn A. Campbell. 1974. Frequency-dependent selection. *Annual Review of Ecology and Systematics* 5.115–38.

Bailey, Charles-James N. 1973. *Variation and linguistic theory*. Washington, DC: Center for Applied Linguistics.

Bailey, Guy. 2002. Real and apparent time. *Handbook of language variation and change*, ed. by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 312–32. Oxford: Blackwell.

Barlow, Michael, and Suzanne Kemmer (ed.) 2000. *Usage-based models of language*. Stanford, CA: CSLI Publications.

Baxter, Gareth J.; Richard A. Blythe; William Croft; and Alan J. McKane. 2006. Utterance selection model of language change. *Physical Review E* 73.046118.

Baxter, Gareth J.; Richard A. Blythe; William Croft; and Alan J. McKane. 2009. Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21.257–96.

Baxter, Gareth J.; Richard A. Blythe; and Alan J. McKane. 2008. Fixation and consensus times on a network: A unified approach. *Physical Review Letters* 101.258701.

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart and Winston.

Boas, Mary L. 1983. *Mathematical methods in the physical sciences*. 2nd edn. New York: Wiley.

Briscoe, Ted. 2000. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language* 76.245–96.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Castellano, Claudio; Santo Fortunato; and Vittorio Loreto. 2009. Statistical physics of social dynamics. *Reviews of Modern Physics* 81.591–646.

Chambers, Jack K. 1995. The Canada-US border as a vanishing isogloss: The evidence of *chesterfield*. *Journal of English Linguistics* 23.155–66.

Chambers, Jack K. 1998. Social embedding of changes in progress. *Journal of English Linguistics* 28.5–36.

Chambers, Jack K. 2002. Patterns of variation including change. *Handbook of language variation and change*, ed. by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 349–72. Oxford: Blackwell.

Chambers, Jack K., and Peter Trudgill. 1998. *Dialectology*. 2nd edn. Cambridge: Cambridge University Press.

Clifford, Peter, and Aiden Sudbury. 1973. A model for spatial conflict. *Biometrika* 60.581–88.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. Harlow: Longman.

Croft, William. 2010. The origins of grammaticalization in the verbalization of experience. *Linguistics* 48.1–48.

Crow, James F., and Motoo Kimura. 1970. *Introduction to population genetics theory*. New York: Harper and Row.

Dawkins, Richard. 1989 [1976]. *The selfish gene*. New edn. New York: Oxford University Press.

de Oliveira, Mário J. 1992. Isotropic majority-vote model on a square lattice. *Journal of Statistical Physics* 66.273–81.

Denison, David. 2003. Logistic and simplistic S-curves. *Motives for language change*, ed. by Raymond Hickey, 54–70. Cambridge: Cambridge University Press.

Devitt, Amy J. 1989. *Standardizing written English: Diffusion in the case of Scotland 1520–1659*. Cambridge: Cambridge University Press.

Ellegård, Alvar. 1953. *The auxiliary* do*: The establishment and regulation of its growth in English*. (Gothenburg studies in English 2.) Stockholm: Almqvist and Wiksell.

Giles, Howard. 1973. Accent mobility: A model and some data. *Anthropological Linguistics* 15.87–105.

Giles, Howard, and Philip M. Smith. 1979. Accommodation theory: Optimal levels of convergence. *Language and social psychology*, ed. by Howard Giles and Robert N. St. Clair, 45–65. Oxford: Blackwell.

Greenberg, Joseph H.; Charles E. Osgood; and Sol Saporta. 1954. Language change. *Psycholinguistics: A survey of theory and research problems*, ed. by Charles E. Osgood and Thomas A. Sebeok, 146–63. Baltimore, MD: Waverly.

Grieve-Smith, Angus B. 2009. *The spread of change in French negation*. Albuquerque: University of New Mexico dissertation.

Harrington, Jonathan; Sallyanne Palethorpe; and Catherine I. Watson. 2000. Does the Queen speak the Queen's English? *Nature* 408.927.

Hartl, Daniel L., and Andrew G. Clark. 2007. *Principles of population genetics*. 4th edn. Sunderland, MA: Sinauer.

Haspelmath, Martin. 1999. Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18.180–205.

Hull, David L. 1988. *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press.

Hull, David L. 2001. *Science and selection: Essays on biological evolution and the philosophy of science*. Cambridge: Cambridge University Press.

Katz, Elihu, and Paul Lazarsfeld. 1955. *Personal influence*. Glencoe, IL: Free Press.

Ke, Jinyun; Tao Gong; and William S.-Y. Wang. 2008. Language change and social networks. *Communications in Computational Physics* 4.935–49.

Keller, Rudi. 1994. *On language change: The invisible hand in language*. London: Routledge. [Translation and expansion of *Sprachwandel: Von der unsichtbaren Hand in der Sprache*, Tübingen: Francke, 1990.]

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1.199–244.

Labov, William. 1966. Hypercorrection in the lower middle class as a factor in linguistic change. *Sociolinguistics*, ed. by William Bright, 84–113. The Hague: Mouton.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors*. Oxford: Blackwell.

Labov, William. 2007. Transmission and diffusion. *Language* 83.344–87.

LePage, Robert B., and Andrée Tabouret-Keller. 1985. *Acts of identity*. Cambridge: Cambridge University Press.

Mazzon, Gabriella. 2004. *A history of English negation*. London: Longman.

Meibauer, Jörg; Anja Guttropf; and Carmen Scherer. 2004. Dynamic aspect of German *-er* nominals: A probe into the interrelation of language change and language acquisition. *Linguistics* 42.155–93.

Milroy, James. 1992. *Linguistic variation and change*. Oxford: Blackwell.

Milroy, James, and Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21.339–84.

Milroy, Lesley. 1987. *Language and social networks*. 2nd edn. Oxford: Blackwell.

MILROY, LESLEY, and JAMES MILROY. 1992. Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21.1–26.

MUFWENE, SALIKOKO. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

MUFWENE, SALIKOKO. 2005. *Créole, écologie sociale, évolution linguistique*. Paris: L'Harmattan.

MUFWENE, SALIKOKO. 2008. *Language evolution: Contact, competition and change*. London: Continuum.

NETTLE, DANIEL. 1999. Using social impact theory to simulate language change. *Lingua* 108.95–117.

NEVALAINEN, TERTTU. 2000. Processes of supralocalisation and the rise of Standard English in the early Modern period. *Generative theory and corpus studies: A dialogue from 10 ICEHL*, ed. by Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg, and C. B. McCully, 327–71. Berlin: Mouton de Gruyter.

NEVALAINEN, TERTTU, and HELENA RAUMOLIN-BRUNBERG. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England*. London: Pearson.

NEVALAINEN, TERTTU; HELENA RAUMOLIN-BRUNBERG; and HEIKKI MANNILA. 2011. The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change* 23.1–43.

NOWAK, MARTIN A. 2006. *Evolutionary dynamics: Exploring the equations of life*. Cambridge, MA: Belknap Press of Harvard University Press.

NIYOGI, PARTHA, and ROBERT C. BERWICK. 1997. Evolutionary consequences of language learning. *Linguistics and Philosophy* 20.697–719.

OGURA, MIEKO, and WILLIAM S.-Y. WANG. 1996. Snowball effect in lexical diffusion: The development of *-s* in the third person singular present indicative in English. *English historical linguistics 1994: Papers from the 8th International Conference on English Historical Linguistics*, ed. by Derek Britton, 119–41. Amsterdam: John Benjamins.

OHALA, JOHN. 1989. Sound change is drawn from a pool of synchronic variation. *Language change: Contributions to the study of its causes*, ed. by Leiv Egil Breivik and Ernst Håkon Jahr, 173–98. Berlin: Mouton de Gruyter.

PIERREHUMBERT, JANET B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 137–57. Amsterdam: John Benjamins.

POPLACK, SHANA, and ELISABETE MALVAR. 2006. Modeling linguistic change: The past and the present of the future in Brazilian Portuguese. *Language variation: European perspectives*, ed. by Frans Hinskens, 169–99. Amsterdam: John Benjamins.

POPLACK, SHANA, and ELISABETE MALVAR. 2007. Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus* 19.121–69.

PRIESTLEY, LEONARD. 1955. Reprise constructions in French. *Archivum Linguisticum* (old series) 7.1–28.

REALI, FLORENCIA, and THOMAS L. GRIFFITHS. 2010. Words as alleles: Connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society B* 277.429–36.

ROGERS, EVERETT M. 1995. *Diffusion of innovations*. 4th edn. New York: Free Press.

ROSENBACH, ANETTE; DIETER STEIN; and LETIZIA VEZZOSI. 2000. On the history of the *s*-genitive. *Generative theory and corpus studies: A dialogue from 10 ICEHL*, ed. by Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg, and C. B. McCully, 183–210. Berlin: Mouton de Gruyter

SANKOFF, GILLIAN, and HÉLÈNE BLONDEAU. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83.566–88.

SAPIR, EDWARD. 1921. *Language*. New York: Harcourt, Brace & Co.

SCHERER, CARMEN. 2007. The role of productivity in word-formation change. *Historical Linguistics 2005: Selected papers from the 17th International Conference on Historical Linguistics*, ed. by Joseph C. Salmons and Shannon Dubenion-Smith, 257–71. Amsterdam: John Benjamins.

STROGATZ, STEVEN H. 1994. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Reading, MA: Addison-Wesley.

STURTEVANT, EDGAR H. 1947. *An introduction to linguistic science*. New Haven, CT: Yale University Press.

TAGLIAMONTE, SALI A., and ALEXANDRA D'ARCY. 2007. Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change* 19.199–217.

TAGLIAMONTE, SALI A., and ALEXANDRA D'ARCY. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85.58–108.

TRUDGILL, PETER. 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* 1.179–95.

TRUDGILL, PETER. 1986. *Dialects in contact*. Oxford: Blackwell.

TRUDGILL, PETER. 2004. *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

TRUDGILL, PETER. 2008. Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society* 37.241–80.

VULANOVIĆ, RELJA, and HARALD BAAYEN. 2007. Fitting the development of periphrastic *do* in all sentence types. *Exact methods in the study of language and text: Dedicated to Gabriel Altmann on the occasion of his 75th birthday*, ed. by Peter Grzybek and Reinhard Köhler, 679–88. Berlin: Mouton de Gruyter.

WANG, WILLIAM S., and CHIN-CHUAN CHENG. 1970. Implementation of phonological change: The Shuāng-fēng Chinese case. *Chicago Linguistic Society* 6.552–59.

WATTS, DUNCAN J., and STEVEN H. STROGATZ. 1998. Collective dynamics of 'small-world' networks. *Nature* 393.440–42.

WEINREICH, URIEL; WILLIAM LABOV; and MARVIN I. HERZOG. 1968. Empirical foundations for a theory of language change. *Directions for historical linguistics*, ed. by Winfred P. Lehmann and Yakov Malkiel, 95–195. Austin: University of Texas Press.

Croft
MSC03 2130, Linguistics
1 University of New Mexico
Albuquerque, NM 87131-0001
[wcroft@unm.edu]

[Received 20 January 2010;
revision invited 9 July 2010;
revision received 1 December 2010;
accepted 14 October 2011]

Blythe
SUPA
School of Physics and Astronomy
Mayfield Road
Edinburgh, EH9 3HZ
United Kingdom
[r.a.blythe@ed.ac.uk]