J. Linguistics **53** (2017), 327–358. © Cambridge University Press 2016
doi:10.1017/S0022226716000141 First published online 7 June 2016

Neutral change[1]

HENRI KAUHANEN

The University of Manchester

(Received 27 April 2015; revised 4 April 2016)

Language change is neutral if the probability of a language learner adopting any given linguistic variant only depends on the frequency of that variant in the learner's environment. Ruling out non-neutral motivations of change, be they sociolinguistic, computational, articulatory or functional, a theory of neutral change insists that at least some instances of language change are essentially due to random drift, demographic noise and the social dynamics of finite populations; consequently, it has remained little investigated in the historical and sociolinguistics literature, which has generally been on the lookout for more substantial causes of change. Indeed, recent computational studies have argued that a neutral mechanism cannot give rise to 'well-behaved' time series of change which would align with historical data, for instance to generate S-curves. In this paper, I point out a methodological shortcoming of those studies and introduce a mathematical model of neutral change which represents the language community as a dynamic, evolving network of speakers. With computer simulations and a quantitative operationalization of what it means for change to be well-behaved, I show that this model exhibits well-behaved neutral change provided that the language community is suitably clusterized. Thus, neutral change is not only possible but is in fact a characteristic emergent property of a class of social networks. From a theoretical point of view, this finding implies that neutral theories of change deserve more (serious) consideration than they have traditionally received in diachronic and variationist linguistics. Methodologically, it urges that if change is to be successfully modelled, some of the traditional idealizing assumptions employed in much mathematical modelling must be done away with.

KEYWORDS: language change, mathematical modelling, neutrality, prestige, S-curves

[1] I thank Ricardo Bermúdez-Otero, David Denison, Tobias Galla and George Walkden for numerous discussions which have contributed greatly to this paper; Laurel MacKenzie, Alan McKane, Mark Muldoon and three anonymous *Journal of Linguistics* reviewers for comments which resulted in important improvements; audiences at the *Student Conference in Complexity Science 2014* (Sussex), the *Manchester Forum in Linguistics 2014* (Manchester), the *International Conference on Computational Social Science 2015* (Aalto University), the *2015 Meeting of the Linguistics Association of Great Britain* (UCL) and the Theory Club of the Cognitive Science Unit at the University of Helsinki, as well as Fernanda Barrientos, Deepthi Gopal, Michaela Hejná and Yuni Kim for feedback; the Faculty of Engineering and Physical Sciences at The University of Manchester for CPU time; and the School of Arts, Languages and Cultures, University of Manchester, and Emil Aaltonen Foundation for financial support.

1. INTRODUCTION

An outstanding problem in diachronic linguistics concerns the extent to which language change is and can be neutral. Once variation arises, are the competing variants created equal, or is change instead motivated by functional, computational, social or other considerations which favour certain variants over others? Expressing an agnostic take on this question, Lass (1997) observes that

> it's perfectly POSSIBLE that both variation and change itself (as a result) are neutral: even selection does not necessarily have to select that which is better 'adapted'. In any case, there are even in biology modes of (apparent) selection that are not (in the Darwinian sense) genuinely 'selective' or 'adaptive' [. . .]. All of these possibilities, given the much better understood nature of variation and change in organisms, need to be considered before any claim for 'function' can be made for either variation or change. (Lass 1997: 354, my emphasis.)

The thrust of this programmatic message, mainly directed at functional explanation but not limited in its scope to explanations of functionalist persuasions, is that the possibility of neutral change has remained, and continues to remain, underinvestigated.

This paper examines that possibility by means of a simple mathematical model of variant competition in a finite population of speakers. Guided by the intuition that language diachrony is typically well-behaved (in a sense to be made precise later), I propose a quantitative metric of well-behavedness and, with the help of computer simulations, investigate how the neutrality hypothesis fares in its light. The upshot of this investigation is that well-behaved neutral change is, indeed, found to be possible if the social network underlying the language community has a suitable topology and dynamics: briefly, if the language community is strongly clusterized, so that it can be partitioned into more central and more peripheral speakers, neutral change is observed. Moreover, it is found that well-behaved neutral change is a consistent, characteristic emergent feature of such social networks: the effect is not a statistical anomaly, but flows naturally and robustly from the way in which the language community is structured and the way in which that structure evolves over time. On the other hand, in a classical well-mixed (unstructured) population we find that change can rarely be neutral AND well-behaved.

The model here studied differs from most previous mathematical models of language change (and, more generally, cultural evolution) in two respects. First, the model does away with the classical idealizing assumption of representing language communities as well-mixed populations, often infinite, and looks instead at finite social networks with non-uniform degree distributions, that is to say networks in which different people have different patterns of connectivity. Second, the model takes into account the fact that human social networks are never static but are constantly being rewired by the removal and addition of individuals: friendship and even family ties are not fixed, people move from one social network to another, and deaths and births occur. The fact that neutral change cannot happen

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

in a classical unstructured population but can happen in populations with suitable topologies and rewiring dynamics points to the need to consider the particularities of socialization in language communities at a level of detail that mathematical and computational models of language change have not attempted so far.

The relevance of the possibility of neutral change to diachronic and variationist linguistics is as follows: unless assumptions of non-neutral motivations of change can be supported for independent reasons, a neutral theory of change remains a viable explanatory strategy. More specifically, the observation that well-behaved neutral change is a characteristic feature of certain kinds of social networks suggests that in some cases of change, neutral selection may be at play in addition to, or instead of, non-neutral selection. Moreover, the parsimonious nature of neutral theory holds promise in clearing up certain puzzles which have traditionally received rather AD HOC solutions in the research literature: by removing the notion of (variant) prestige, a neutral mechanism can provide a fresh, bias-free sociolinguistic take on change, as I will argue in Section 6.

2. NEUTRALITY

The possibility that language change might be neutral has, traditionally, received little attention in historical and variationist linguistics. Aside from occasional remarks such as Lass's (1997) cited above, and Postal's (1968: 283–285) suggestion that language change is random, non-motivated 'fashion change', the neutrality hypothesis has received serious consideration mainly from Trudgill (2008), who, in questioning the role of identity in new-dialect formation, suggests that dialect contact and dialect mixing work in 'automatic', non-biased ways. Although Trudgill's position is avowedly anti-identity, and thus rejects one form of (social) bias, it is not entirely clear whether his account might not admit some other form of bias, however. In fact, whether argued for or against, the neutrality hypothesis is rarely defined in precise, unequivocal terms in the literature. In this section, therefore, I will explicate the hypothesis by putting forward a definition of neutrality and contrasting neutral change with change governed by non-neutral factors.

Throughout this paper, I will focus on a situation in which a fixed number of linguistic variants are in competition in a specific linguistic domain. To keep the discussion maximally general, I will not make further assumptions about the nature or composition of these variants. Depending on the application, a variant could be a complete parametric specification of Universal Grammar, a single value of one particular parameter, an allophone of a phoneme, and so on. What matters is that there are a number of variants, each of which COULD be adopted, in principle, by any speaker. The neutrality hypothesis can then be stated, in intuitive terms, as follows.

(1) *Neutrality*
The probability with which a speaker acquires a certain linguistic variant out of a number of competing variants equals the relative frequency of that

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

> variant in the speaker's neighbourhood, MODULO a small probability of innovating, uniformly at random, another variant from among all (biologically, cognitively) possible variants.

'Neighbourhood' here means the speaker's linguistic neighbourhood, a term to be explicated in more detail shortly. The content of the neutrality hypothesis, then, is that variant selection in individual speakers is controlled by the frequencies of the competing linguistic variants: apart from a small amount of random noise that accounts for variant innovation, no considerations other than the frequency distribution of variants could affect which variant an individual acquires or adopts. To borrow terminology from biology, under an assumption like (1), change (evolution) is FREQUENCY-DRIVEN, and the competing variants do not have differential FITNESSES which could bias the process of adoption, favouring one variant over another. In contrast to fitness-driven (e.g. Darwinian) selection, in neutral change the adoption of a given variant is not adaptive in any sense. The innovation events, when they do occur, are likewise neutral: the innovatory variant is chosen from among all possible variants uniformly at random, so there is no bias towards innovating any particular variant.[1]

It will be instructive at this point to briefly consider models and modes of explanation in diachronic linguistics which are either explicitly or implicitly NON-neutral, so as to bring the contrast into sharper relief. The typological-functional explanations Lass (1997) alludes to in the quotation cited in Section 1 form an obvious but important instance: there, it is assumed that a linguistic variant can perform better or worse in any given role; linguistic forms perform communicative, cognitive and other functions, and some do this better than others (e.g. Anttila 1989). Under a functionalist hypothesis, processes of change will then be guided by people's intuitions (conscious or subconscious) concerning the performance of different variants in serving these various functions. Change is not neutral, since adopting some variants is deemed, in one sense or another, better than adopting other variants, to the extent even that possible language states are classified into 'consistent' and 'transitional', or 'preferred' and 'dispreferred' ones (e.g. Hawkins 1990, Vennemann 1993; also see the critical discussion in Lightfoot 1999: 85–87 and PASSIM). In the most extreme versions of this

[1] Although one should beware of drawing facile cross-disciplinary analogies, it is worthwhile to point out that neutral mechanisms of change have been proposed in evolutionary biology. In biological evolution, a variant (a genotype or a phenotype, or some part of one) is said to be selectively neutral or simply neutral if having that variant confers neither a selective advantage nor a selective disadvantage. Depending on one's take on the level of selection debate (Reeve & Keller 1999), this implies that a neutral variant will neither increase nor decrease the fitness of its bearer, of the bearer's species or of that variant itself. This mechanism of neutral evolution (Alonso et al. 2006) is to be contrasted with Darwinian natural selection, which operates on complicated fitness landscapes that confer selectional pressures on the competing replicators or vehicles. Although (non-neutral) natural selection remains the DE FACTO mechanism for explaining evolution on various levels of biological organization, neutral theories have been proposed and defended for molecular evolution (Kimura 1994) as well as in ecology for competition within a trophic level (Hubbell 2001).

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

framework, linguistic systems are viewed as teleological (Itkonen 1981, 1982), and 'language change is language improvement' (Vennemann 1993: 322).

Another way of flouting the neutrality hypothesis (1) is by way of considerations of economy of computation or of production. Starting with Lightfoot's (1979) Transparency Principle, the diachronic generative syntax literature has generally favoured explanatory frameworks of this kind, where innate principles or third-factor processing constraints are taken to bias the acquisition of syntax. To take a more recent example, Roberts & Roussou (2003) assume a Merge-over-Move principle to account for parametric reanalysis and grammaticalization under the right kind of trigger experience. In much the same vein, theories and models of sound change which appeal to articulatory (e.g. aerodynamic, inertial) constraints as a motivation or cause of change are non-neutral, in that speakers are assumed to be biased to produce certain (e.g. centralized, lenited) phonetic variants over other, possible ones. Such constraints are said to give rise to variation in the auditory input available to the listener, eventually causing change in the speaker–listener loop (Ohala 1989, Pierrehumbert 2001).[2]

A third mechanism for non-neutral change is constituted by different kinds of social biases. In a typical prestige-based explanation, for example, speakers are said to accommodate towards variants they consider prestigious or associate with a particular social group (Labov 1972). Again, the acquisition or adoption of a variant under such circumstances will be non-neutral, as it is not driven simply by the frequency distribution of variants in the speakers' environments. In a prestige-based explanation, the explanatory onus is on speakers' estimations of the social 'desirability' of particular variants, either overt or covert, and even sociolinguists who are careful to consider other components of processes of actuation and propagation, such as variation in social network structure, have usually assumed (often A PRIORI) that at least a small amount of prestige is necessary for innovatory forms to propagate through a language community. Thus,

> in view of the very general finding of sociolinguistic research that the prestige values attached to language are often quite covert and difficult to tap directly, we may suggest that a successful innovation needs to be evaluated positively, either overtly or covertly. This is OF COURSE a necessary but not a sufficient condition for its ultimate adoption. (Milroy & Milroy 1985: 368, my emphasis.)

These non-neutral mechanisms have been implemented, in varying degrees of detail, in computational models of language change. For instance, Ke, Gong & Wang (2008) find that an innovatory variant must in their model be biased over the prevailing conventional variant, sometimes twentyfold, in order to secure successful propagation. Similarly, in a model of sociolinguistic factors in change, Fagyal et al. (2010) find that speakers must be biased to adopt variants from

[2] It perhaps needs to be stressed in this connection that the point of contest between neutral and non-neutral theory is not whether things such as computational or articulatory constraints EXIST, but whether they are OPERATIVE or CAUSATIVE in language change on a population level.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

speakers who are both well-connected and prestigious in order for the model to generate propagation curves that have the broad outline of an S-curve, the S-curve being taken as a basic desideratum that a model of change should be able to replicate. Finally, in what is perhaps the most extensive computational study so far of the effect of various kinds of biases in variant adoption and propagation, Blythe & Croft (2012) find that in their extension of the utterance selection model of language change (Baxter et al. 2006, 2009), a neutral, non-biased mechanism is unable to generate realistic time series of change such as S-curves.

The framework adopted in the latter study deserves more detailed comment, as it provides a useful sociolinguistic taxonomy of selection mechanisms in language change more generally. Leaning on Croft's (2000) evolutionary theory of language change, itself based on Hull's (1988) general theory of selection processes, Blythe & Croft (2012: 272–277) classify replication mechanisms into four categories: (i) neutral evolution, which is random, frequency-driven drift; (ii) neutral interactor selection, in which speaker–speaker interaction frequencies play a role; (iii) weighted interactor selection, in which interactions between different speakers are weighted differently; and (iv) replicator selection, in which the competing linguistic variants themselves are weighted differently. The neutrality hypothesis (1), as here defined, corresponds to (i) and (ii) with the proviso that interaction frequencies play a role at the point of acquisition, not (necessarily) across the lifespan of speakers as in the usage-based model of Blythe & Croft (2012). The difference between the neutral mechanisms (i)–(ii) and the non-neutral ones (iii)–(iv) is that in the former case no social evaluations take place, whereas in the latter case either speakers or the competing linguistic variants themselves receive potentially differential evaluations, which fact is taken to be the motor of change. In fact, Blythe & Croft (2012) find that, in their model, S-curves are reliably obtained only for replicator selection, in other words for selection where the social evaluations mark linguistic variants directly, in the Labovian sense.

Although the aforementioned formal models have elucidated important aspects of variation and change in natural language, they have their limitations. Most importantly from the point of view of our present concerns, each of the three models mentioned above (Ke et al. 2008, Fagyal et al. 2010, Blythe & Croft 2012) represents language communities as static networks of speakers. That is to say, even though these models are rich enough to represent differences in social network topology, or differences and asymmetries in the probabilities with which different speakers interact, they lack a mechanism for evolving that topology, and consequently fail to model the social DYNAMICS of a language community: in the aforementioned models, there is no way for individual speakers to be removed from or added to the network, or for their connection sets or interaction probabilities to evolve within a single simulation run. Importantly, the models then fail to address the question of whether and how that social dynamics might affect the linguistic variant dynamics operating on the social network. The assumption of a static network clearly does not hold for human societies – and the longer the time spans of any particular changes we may be interested in explaining, the worse this approximation becomes. Moreover, recent research in fields such as

mathematical epidemiology and evolutionary game theory has demonstrated that the qualitative features of a dynamical system operating on a network may be significantly altered when that underlying network is endowed with a dynamics of its own (Gross et al. 2006, Traulsen, Santos & Pacheco 2009). It is then reasonable to ask whether previous computational studies of language change may not have arrived at the wrong conclusions concerning the neutrality hypothesis by making the wrong kinds of idealizing assumptions.

3. MODEL

In this section, I will define the present model in intuitive terms, using as little mathematical notation as possible. A technical, mathematical definition can be found in Appendix A.

As outlined in Section 2, I will be focussing on a situation in which some number C of variants are in competition; often, the focus is on cases where C is small, but in general this number could be arbitrarily large. The variants are assumed to be distributed across a language community in the following sense: each one of N speakers will entertain exactly one of the C variants at any time. The speakers themselves are distributed on a social network, and the connections a speaker has in this network will affect their process of variant acquisition or adoption. For simplicity, I assume that the network is symmetric, that connections are binary and that the network is not multiplex. In other words, if speaker i is connected to speaker j, then speaker j will also be connected to speaker i; each pair of speakers is either connected or not connected (there is no notion of 'weight' of connection); and only one connection is allowed between any two speakers.

The set of speakers to whom a given speaker is connected I shall call the NEIGHBOURHOOD of that speaker; using basic graph-theoretical terminology, the cardinality of this set is called the speaker's DEGREE (in other words, the degree of a speaker is simply the number of speakers that speaker is connected to). When new speakers acquire their variant, their neighbourhoods are all important. In line with (1), I assume that the probability of acquiring variant r ($r = 1, \ldots, C$) equals the relative frequency of that variant in the speaker's neighbourhood, MODULO random noise which is taken to model processes of innovation. This random noise is inserted into the model as an INNOVATION PARAMETER μ, which ranges from 0 to 1 and gives the probability that the speaker picks a variant from among the C possible variants uniformly at random. It is clear that this probability has to be rather low for language communities to display a degree of coherence in what variants they use – and this expectation is borne out by the simulations reported in Section 5 below. The parameter is, however, an essential part of the model, for without it, variation could not arise in the first place.[3]

[3] To see this, suppose that each individual in the community happens to use the same variant, so that the relative frequency of this variant in the community equals 1. If $\mu = 0$, then, in line with (1), any new speaker inserted into the network will acquire the said variant with probability 1, and change is impossible.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

To model social dynamics, the network of speakers is mixed by a graph-rewiring process over time, as follows. At each iteration step, one of the speakers is selected for removal uniformly at random and is replaced by a new speaker, whose social connections are set according to a socialization algorithm. The new speaker then acquires their variant as outlined above. The socialization algorithm is modelled on the intuition that human social networks normally contain both more and less connected individuals (cf. Barabási & Albert 1999), and operates as follows. Let σ be a real number with $0 \leqslant \sigma \leqslant 1$ and K an integer with $1 \leqslant K \leqslant N - 1$. The new speaker is then given exactly K connections according to the following procedure. For each connection, the speakers in the network are first rank-ordered into a queue in terms of decreasing degree in such a way that the order of speakers having the same degree is random, but speakers with higher degree occur earlier in the queue than speakers with lower degree. Then, with probability σ, a connection is made to the first speaker in this queue, and with the remaining probability mass $1 - \sigma$ a connection is made to a speaker chosen uniformly at random from the queue. Once the connection is established, this speaker is removed from the queue and the procedure is iterated until the new speaker has received K connections. (Note that this does not imply that each speaker will, at any point of time, have exactly K connections: speakers' degrees will change during their lifetimes thanks to the graph-rewiring process, as other speakers are removed from the network and replaced by new ones.) Different values of σ, a PREFERENTIALITY PARAMETER, then give rise to networks with different amounts of clusterization around a central component, and different combinations of K and σ can be used to model different kinds of population structures: for high K and low σ ($K \approx N - 1$ and $\sigma \approx 0$), the population is well-mixing, whereas for small K and high σ, for instance, the network has a star-like appearance, with a clear partitioning into central and peripheral individuals (Figure 1).

The model assumes invariant and categorical speakers – speakers who fix onto one of the competing variants at the point of acquisition and never change thereafter – this assumption being made in the interest of computational and mathematical tractability. Although some linguistic features are known to remain variable throughout a speaker's lifetime (e.g. Harrington 2006, Sankoff & Blondeau 2007), there is, equally, evidence that for other features late-life change is unlikely or outright impossible. For instance, a number of studies have shown the existence of 'hard features' – features that only young children manage to acquire and for which plasticity is lost as the speaker matures (these include, for example, phonological features with lexically irregular conditioning; see Kerswill 1996 for a review of a number of relevant studies). Moreover, there is evidence that early categoricity predicts late-life stability: in a longitudinal panel study of a dozen phonetic variables undergoing change in a rural Finnish-speaking community, Nahkola & Saanilahti (2004) found significant late-life change only in speakers who had acquired features as variable ones. For categorical or near-categorical features, late-life change appears to be unlikely. These findings suggest that

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

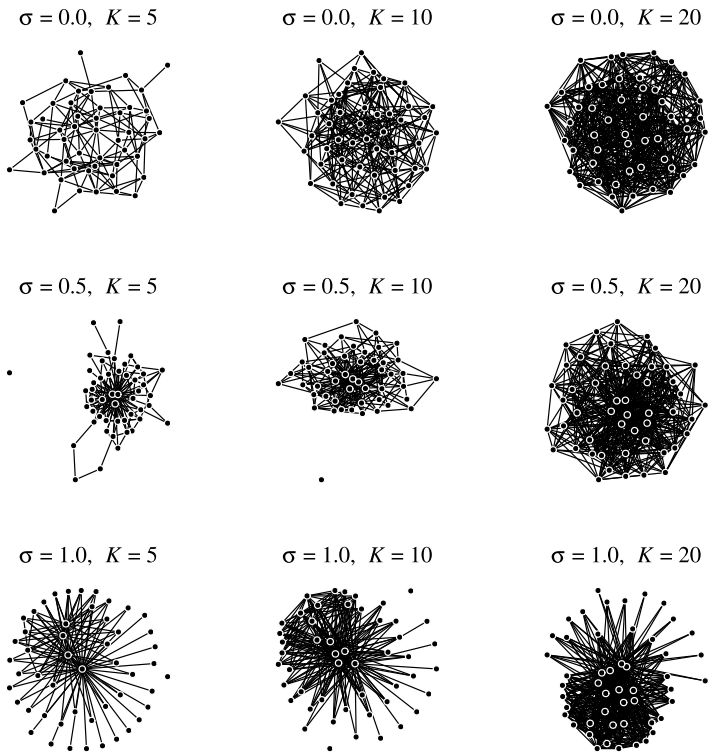

Figure 1
Different values of the preferentiality parameter σ, combined with varying values of K, lead to networks with different amounts of clusterization. Note that the networks are not static but are rewired over time by the removal and addition of speakers; as a consequence, individual speakers may at times become disconnected from the rest of the network. For the networks in this figure, $N = 50$.

categorical, acquisition-driven change is one way in which languages change, and that the assumption of invariant speakers is therefore not unduly unrealistic – but future modelling work should, of course, investigate the consequences of relaxing the assumption.

4. WELL-BEHAVING

Evaluation of the neutrality hypothesis (1) requires us to compare the output of the neutral model defined in Section 3 against some sort of standard. More specifically, our interest is in two questions. (i) Does the neutral mechanism give change in the first place? (ii) If so, do the trajectories of change look anything like real life change trajectories? In this section, I introduce a way of operationalizing

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

these two questions by way of a notion of the 'well-behavedness' of change. Following a preliminary, intuitive characterization, I will show how this notion can be formalized in mathematical, quantitative terms, so that the presence or absence of well-behaved change can be detected in simulation data generated by mathematical models of language change. The following section then proceeds to evaluate the current model vis-à-vis this quantitative operationalization in order to investigate the viability of neutral change.

Change, in a very general sense, can be said to occur when the distribution of linguistic variants over a language community changes. In most cases of interest to the historical linguist, such changes proceed from a state where one variant is DOMINANT or nearly so in the community (has relative frequency of, or close to, 1) to another such state where another variant has become (near) dominant. Moreover, when diachronic data are consulted, such SHIFTS between two dominance states are, as a general rule, found to be remarkably smooth, or well-behaved. Language change is not a random walk in the frequency space of possible linguistic variants; on the contrary, time series of changes can often be approximated to a good degree using a sigmoid, or S-shaped, function (Bailey 1973, Kroch 1989, Croft 2000, Blythe & Croft 2012). Although it remains unknown whether all changes follow an S-curve, and, if so, whether the detailed 'shape of S' is the same in all changes (Niyogi & Berwick 1997, Denison 2003, Ghanbarnejad et al. 2014), propagation curves of linguistic variants tend to be reasonably MONOTONE: they are unlikely to show oscillations by repeatedly inflecting up and down in the time domain, but rather proceed smoothly from one dominance state to another.

With these considerations in mind, I suggest that any model of language change should fulfil the following three criteria, which I here lay out, in a programmatic manner, as characteristic properties of language diachrony under normal conditions.

(2) *Dominance*
For most of the time, the language community relaxes into a state in which one variant is (nearly) dominant, so that most or even all speakers use that variant.

(3) *Shifting*
Upon introduction of an innovatory variant, this innovation may nonetheless begin to spread and eventually permeate (most of) the community; thus, the community may shift from a state of dominance by variant r to a state of dominance by variant $r' \neq r$.

(4) *Monotonicity*
Such shifts proceed in a monotone manner, with the frequency of the invading variant increasing and the frequency of the receding variant decreasing along smooth propagation curves.

A language community that fulfils all three criteria I shall call WELL-BEHAVED.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

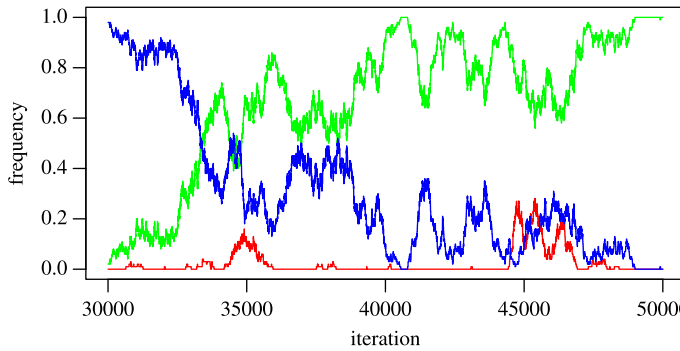

Figure 2
Portion of an ill-behaved history that violates dominance and monotonicity in a system of three variants. This trajectory was generated with parameter settings $N = 100$, $K = 10$, $C = 3$, $\mu = 0.005$ and $\sigma = 0$.

(5) *Well-behavedness*

Language change is well-behaved if, and only if, it satisfies dominance, shifting and monotonicity.

To fix these ideas, let us inspect two simulation histories generated by the model defined in Section 3 qualitatively. Figure 2 shows a snapshot of an ill-behaved history in a three-variant system violating both dominance and monotonicity; this history was generated by setting $\sigma = 0$ and $\mu = 0.005$, the remaining parameters having the values $N = 100$, $K = 10$ and $C = 3$. Although the language community does display a kind of change, and hence exhibits shifting, this change is not monotone: there is too much zig-zagging movement in the propagation curves of the individual variants for this trajectory to be considered well-behaved. Moreover, the community does not settle on a dominant variant for any extended period of time. The history in Figure 3, on the other hand, illustrates an entirely different situation, even though it was produced by the very same neutral mechanism. This history, generated with $\sigma = 1$, the other parameter values remaining the same, satisfies all three conditions: dominance, shifting and monotonicity.

Well-behaved neutral change is, then, possible. It remains to show that this is not merely a chance occurrence but a consistent behavioural characteristic of the model for certain ranges of model parameter values. For this, a quantitative analogue of each of the criteria (2)–(4) is needed, one that can be calculated over a large batch of simulation runs for a number of possible combinations of model parameter values in order to estimate, in a statistically robust manner, to what extent that criterion is satisfied by that combination of model parameters. A formal definition of such quantitative measures is given in Appendix B; here, I introduce the measures in prose.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

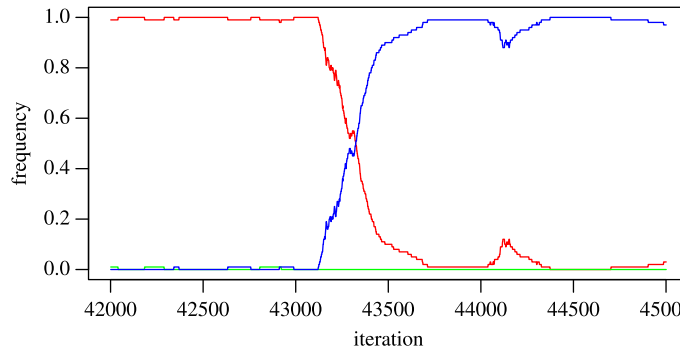

Figure 3
Portion of a well-behaved history satisfying dominance, shifting and monotonicity. For this simulation, $N = 100$, $K = 10$, $C = 3$, $\mu = 0.005$ and $\sigma = 1$.

To estimate to what extent a given simulation run satisfies dominance, I shall use a measure of dominance time D_δ that ranges from 0 (language community never dominant) to 1 (language community dominant all the time). The parameter δ, a small real number, controls how strictly dominance is to be measured. More precisely, for given δ, I say that a language community is δ-DOMINANT if some one variant has a relative frequency equal to or greater than $1 - \delta$; the D_δ measure then gives the proportion of time (in relation to the length of the entire simulation run) the system spends in a state of δ-dominance. Variation of δ allows one to calculate dominance times to varying degrees and, INTER ALIA, to subsume a notion of stable variation under the notion of dominance. For example, setting $\delta = 0.3$, we would call a community dominant if one of the competing variants had a relative frequency of at least 0.7 – allowing the rest of the frequency mass, a number bounded from above by 0.3, to be distributed among the remaining variants in any manner.

To measure shifting, I shall simply determine, for a given simulation run, the number of times the community shifts from a state of δ-dominance by some variant r to a state of δ-dominance by another variant $r' \neq r$, for a predefined dominance level δ. In what follows, I shall denote this shifting measure with S_δ, where δ gives the desired dominance level.

Finally, to quantify monotonicity, I shall look at the autocorrelation properties of individual simulation histories. The idea is to place a short time window of some τ time steps at a specific time point t_0 of a history (so that the time window extends from t_0 to $t_0 + \tau$), and then to count how many times the frequency of each competing variant both increases and decreases inside that window. To be more precise, let $m_r^+ = m_r^+(t_0, \tau)$ be the number of times the frequency of variant r increases within such a time window, and let $m_r^- = m_r^-(t_0, \tau)$ be the corresponding count of decreases. It is then easy to see that the product $m_r^+ m_r^-$

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

equals 0 if, and only if, the frequency curve of variant r is monotone in the window: the product is zero if and only if at least one of the multiplicands is zero, which is equivalent to monotonicity. For technical reasons explained in Appendix B, I next take the square root of this product and sum over competing variants, arriving at $\sum_{r=1}^{C} \sqrt{m_r^+ m_r^-}$. Finally, an average is taken over different selections of the time window start point t_0 (in other words, the time window is slid across the entire simulation history), and an inversion and a normalization are performed so that the final measure, M_τ, ranges from 0 in the NON-monotone case to 1 in the perfectly monotone case, with τ controlling the resolution at which monotonicity is measured.

Some of the properties of this measure, proved in Appendix B, are worth mentioning here: it can be shown that $M_\tau = 0$ for any even[4] integer τ if the frequency of at least one variant zig-zags persistently, increasing at every other iteration step and decreasing at every other; that $M_\tau = 1/3$ for large τ if the history is a random walk (so that for any variant and any iteration step, it is equally probable that the frequency of this variant increases, decreases or stays the same at that step); and that $M_\tau = 1$ for any τ if the frequency curve of each variant is monotone in all windows of size τ. Figure 4 illustrates a few histories with their corresponding M_τ scores for different window sizes τ, to give an idea of what amount of smoothness to expect for individual trajectories, given a value of M_τ.

We then have three measures, D_δ, S_δ and M_τ, to measure dominance, shifting and monotonicity in individual simulation runs. Before proceeding to an application of these measures to the neutral model, it is perhaps in order to clarify the purpose of defining the measures in the first place. Importantly, the above operationalization of the well-behavedness of linguistic change should not be taken to imply that no language community is ever ILL-behaved. After all, empirically demonstrated cases exist of both stable variation (violation of strict dominance; cf. Wallenberg 2013) and zig-zagging or failing changes (violation of monotonicity; Postma 2010, Coussé & De Sutter 2012). The purpose of introducing dominance, shifting and monotonicity as conditions of well-behaved change is to fashion a litmus test for the neutral model: insofar as the model satisfies the three conditions, it can be taken seriously as a mathematical model of language change. With the above operationalization, dominance and monotonicity are actually continuous quantities, ranging from 0 to 1, and thus admit a notion of degree. Requiring the neutral model to satisfy well-behavedness to a large degree is the strictest possible analytical test to which the model can be subjected in this regard, and if real life language communities are found to be less well-behaved than this, then the case for neutral change is correspondingly strengthened.

[4] The technical reason for the restriction here, without loss of generality, to even (rather than odd) integers is explained in Appendix B.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

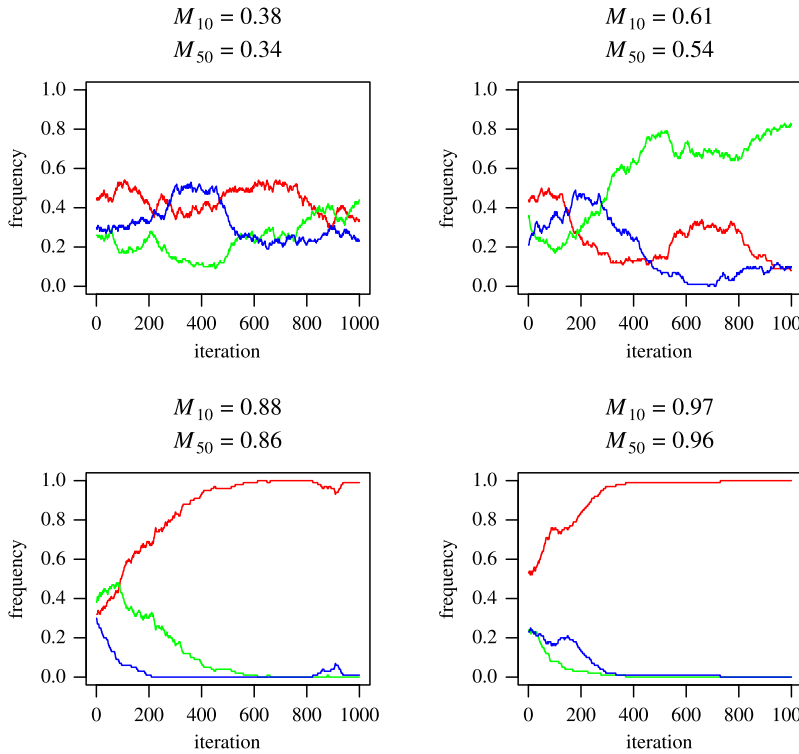

Figure 4
Four histories with their corresponding monotonicity scores M_τ for two different window sizes $\tau = 10, 50$. Note that for a random walk the expected value of M_τ is $1/3$ (see text), and that M_τ approaches 1 as the history becomes more and more monotone.

5. SIMULATIONS

To investigate to what extent the model defined in Section 3 satisfies dominance, shifting and monotonicity, or criteria (2)–(4), a number of computer simulations of the model were run using a range of model parameter settings. For each combination of model parameters investigated, 50 simulations were run to arrive at the averages reported below, and in each simulation, a social network of $N = 100$ speakers was assumed. The simulations were run in parallel on a high-throughput computing cluster, with the pseudorandom number generator seeded using environmental noise to ensure statistical independence of simulation runs. Before starting each actual linguistic simulation, the social network algorithm was iterated for $100N = 10^4$ iterations so that the degree distribution of the network settled; each actual linguistic simulation (apart from the simulations reported in Section 5.5; see below) lasted for 5×10^4 iterations and started from a state in which one of the competing variants had strict dominance (relative frequency 1, or, in other words, δ-dominance with $\delta = 0$).

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

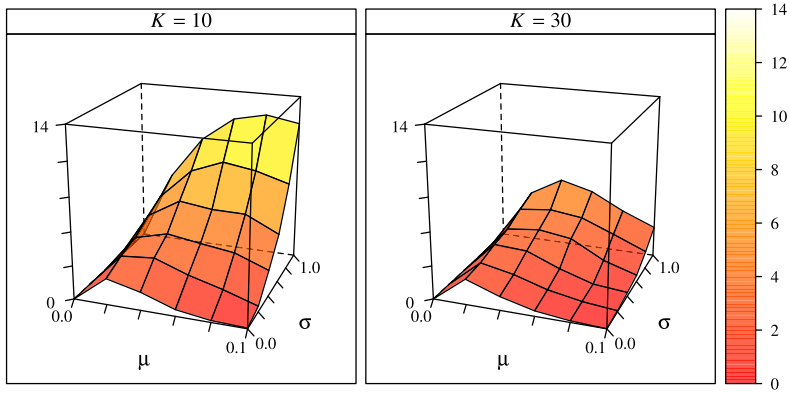

Figure 5
Shifting $S_{0.1}$ in a system of $C = 3$ competing variants, for various values of preferentiality σ and innovation rate μ and for attachment set sizes $K = 10, 30$, calculated using a dominance threshold of $\delta = 0.1$; averages over 50 simulation runs. Neutral change is supported the best by tightly clusterized communities (small K, high σ).

5.1 *Main result*

Figure 5 gives shifting scores for a system of $C = 3$ competing variants, for various values of preferentiality σ and innovation rate μ, and for two different values of attachment set size K, using a dominance threshold of $\delta = 0.1$. The results indicate that each of these model parameters has its effect on shifting ability: keeping K and μ constant, the effect of increasing σ from 0 towards 1 is a monotonic increase in shifting; for μ, on the other hand, an optimal value exists that supports shifting ability the best. Increasing K, in turn, has the effect of flattening the shifting measure with respect to σ: as the population becomes more and more well-mixing, preferential connectivity naturally ceases to have an effect and shifting becomes rarer. In sum, change is the more probable the smaller K is and the larger σ is – the more clusterized the community is around a central component (cf. Figure 1) – provided that innovations (μ) occur at a suitable rate.

Figure 6 plots dominance times and monotonicity, the former calculated assuming $\delta = 0.1$, the latter computed using a window size of $\tau = 10$; variation in τ has only a minor effect on the monotonicity measure (not reported; but cf. Figure 4). The main finding with respect to dominance is that increasing the innovation rate μ results in a sharp drop in this measure, with the value of σ attenuating the effect a little, so that communities with higher preferentiality σ remain dominant for larger μ than communities with lower preferentiality σ. A similar, but much less drastic, drop as a response to variation in μ is observed for monotonicity.

To gauge what combinations of model parameter values support well-behaved neutral change the best overall, we can consider the product of the three measures,

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

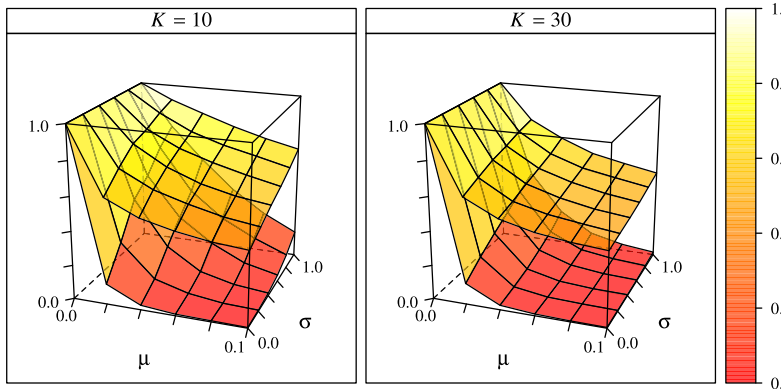

Figure 6
Dominance $D_{0.1}$ (bottom surface) and monotonicity M_{10} (top surface) in a system of $C = 3$ competing variants; averages over 50 simulation runs. Both dominance and monotonicity drop as the innovation rate μ is increased, with large preferentialities σ attenuating this effect.

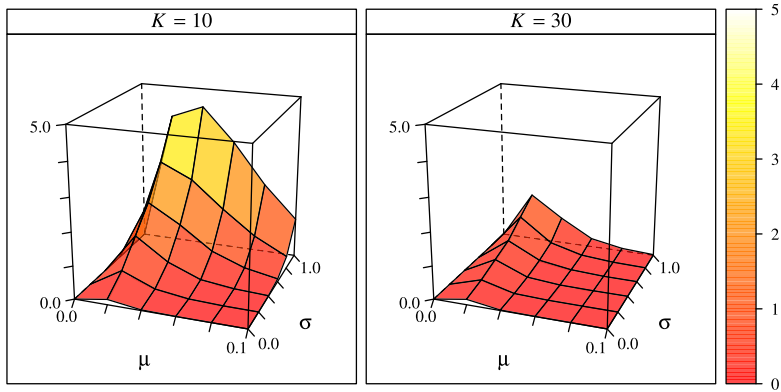

Figure 7
The combined well-behavedness measure $S_{0.1}D_{0.1}M_{10}$ for a system of $C = 3$ variants; averages over 50 simulation runs. Overall, well-behaved neutral change is supported best by tightly clusterized language communities and by innovation rates μ that are low but not too low.

namely $S_{0.1}D_{0.1}M_{10}$. Figure 7 gives this product, and we find that communities with low K, high σ and intermediate μ are the most likely to exhibit well-behaved neutral change.

5.2 *Effect of number of variants*

In the above simulations, the number of competing variants was fixed at $C = 3$. This is a rather small number, and it is reasonable to ask whether the behaviour

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

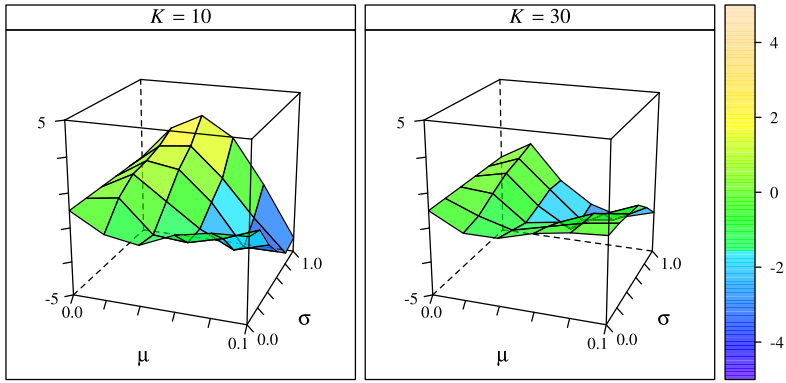

Figure 8
Difference ($B - A$) in shifting $S_{0.1}$ between (A) the 3-variant system of Section 5.1 (Figure 5) and (B) another system with $C = 30$ competing variants CETERIS PARIBUS. For large σ, the 30-variant community shifts more than the 3-variant system if μ has a modest value; for larger μ, the reverse obtains.

of the system would change if more variants were available to speakers. To investigate this, another batch of simulations was run using identical model parameter settings except that the number of competing variants was now fixed at $C = 30$.

Increasing the number of variants turns out to have a non-trivial effect on the well-behavedness of a neutral system. Figure 8 gives the difference between the shifting scores received by the new batch of simulations and those received by the simulations of Section 5.1. Here, we find that for certain combinations of preferentiality σ and innovation rate μ, the community with $C = 30$ shifts more than the community with $C = 3$, whereas for other model parameter combinations the reverse is true: increasing the number of competing variants improves shifting for large σ, but only if μ has a modest value.

Figure 9 reports, similarly, the difference in dominance and monotonicity scores received by the two batches of simulation runs. Increasing C leads to slightly lower dominance and monotonicity overall, an effect that is the strongest for an intermediate range of values of μ.

Thus, overall, allowing speakers a larger space of grammatical options can have the effect of increasing the probability of change, but only at the cost of some reduction in how well-behaved that change is in terms of dominance and monotonicity.

5.3 *Effect of dominance threshold*

The dominance threshold $\delta = 0.1$ used above is rather strict: it demands a variant to have a relative frequency of more than 0.9 in order for that variant to be considered dominant. Lowering the dominance threshold is expected to increase

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

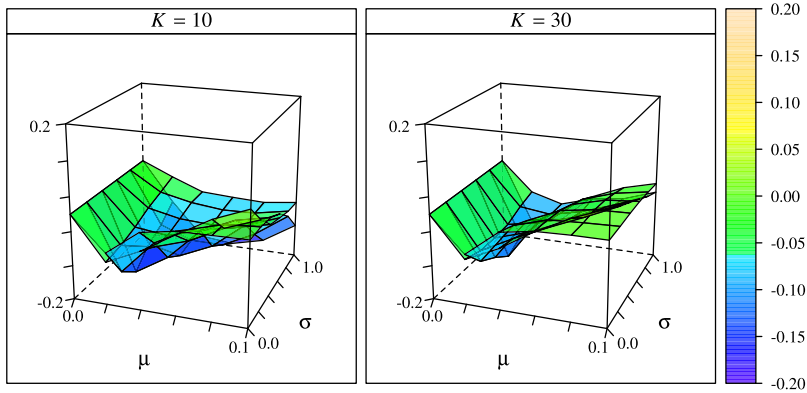

Figure 9
Difference $(B - A)$ in dominance $D_{0.1}$ and monotonicity M_{10} between (A) the 3-variant system (Figure 6) and (B) a system with $C = 30$ competing variants CETERIS PARIBUS. Increasing the number of competing variants leads to slightly lower dominance and monotonicity overall, the effect being the most pronounced for intermediate values of the innovation rate μ.

both shifting and dominance, and this expectation is confirmed by calculations of S_δ and D_δ using a less stringent dominance threshold of $\delta = 0.3$ (Figures 10 and 11). A non-trivial finding is that the preferentiality parameter σ has a strong effect on dominance for less extreme dominance thresholds: for $\delta = 0.3$ and $K = 10$, for instance, $\sigma = 0$ implies practically no dominance if μ is of the order of 0.1, while for $\sigma = 1$ dominance times remain in the > 0.5 region for such innovation rates. Thus, the model predicts that when change is neutral, stable variation is supported best by language communities that are tightly clusterized.

5.4 *Effect of rewiring dynamics*

We can also ask whether it is just the topology of the social network that licenses well-behaved neutral change for certain ranges of parameter values, or whether the social dynamics induced by the removal and addition of speakers plays a role. To investigate this, another batch of simulations was run with parameter settings identical to those of the first ensemble (Section 5.1), but with the rewiring dynamics turned off. (In other words, the network was first rewired for 10^4 iteration steps, as above, to give it the topology induced by the particular choice of K and σ in each case, so that the network had the same topology as in the rewired case. However, the rewiring dynamics was turned off at this point, so that during the actual linguistic simulation no rewirings took place and the network was thus static.) Figure 12 gives the difference in the overall well-behavedness score – the product $S_{0.1} D_{0.1} M_{10}$ – between these two ensembles. For less clusterized networks (large K or small σ) the difference is negligible, as would be expected. For strongly clusterized networks, however, an entirely different picture emerges

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

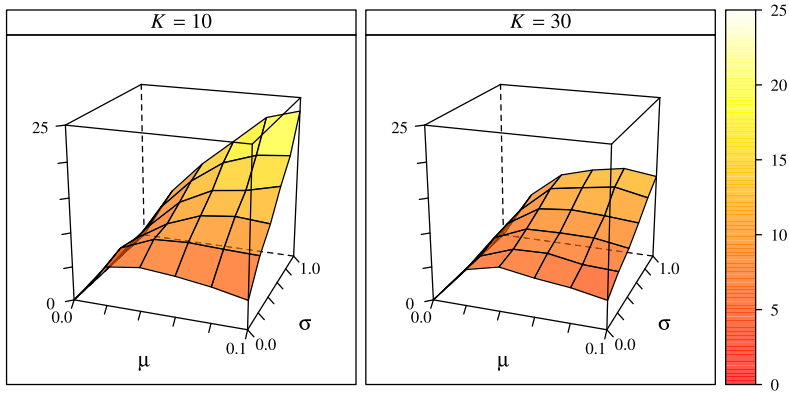

Figure 10
Shifting $S_{0.3}$ for a system with model parameter values identical to those of the system of Figure 5, calculated using a less stringent dominance threshold of $\delta = 0.3$. Adjusting the threshold in this way leads to more shifting events across all of the model parameter space.

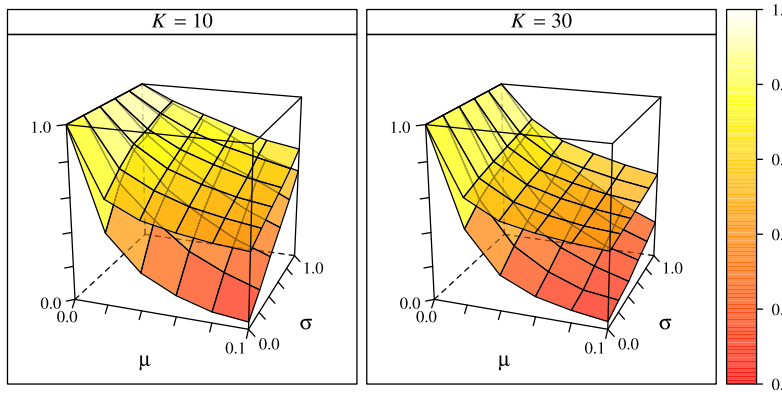

Figure 11
Dominance $D_{0.3}$ (bottom surface) and monotonicity M_{10} (top surface) for a system with model parameter values identical to those of the system of Section 5.1 (cf. Figure 6), calculated using a less stringent dominance threshold of $\delta = 0.3$. For this laxer dominance threshold, network preferentiality σ has a strong effect on dominance: $\sigma = 0$ implies essentially no dominance if innovations occur at a rate of about $\mu = 0.1$, whereas for more tightly clusterized communities ($\sigma \approx 1$) dominance times remain in the > 0.5 region for such innovation rates. This means that stable variation – δ-dominance with a lax dominance threshold such as $\delta = 0.3$ – is supported best by language communities that are tightly clusterized, when change is neutral.

when the rewiring dynamics is removed: the community without rewiring displays consistently lower well-behavedness scores.

This finding may appear puzzling at first sight, but is actually connected in a natural way to one of the central idealizing assumptions of the model,

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

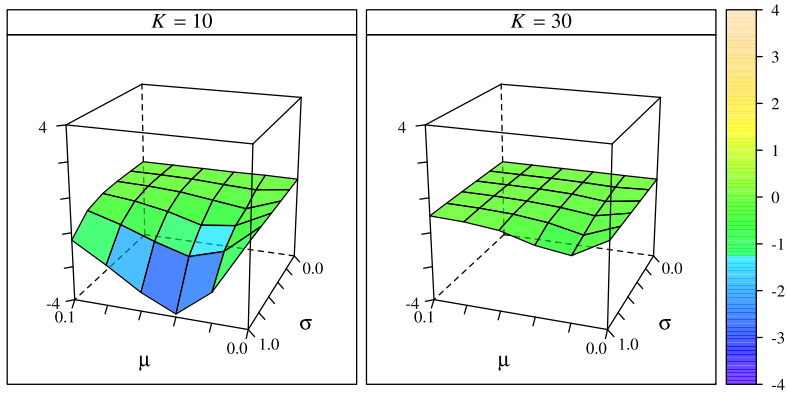

Figure 12
Difference $(B - A)$ in the overall measure of well-behavedness, $S_{0.1}D_{0.1}M_{10}$, between (A) the system of Section 5.1 (Figure 7) and (B) another one with the rewiring dynamics turned off, model parameter settings remaining the same. When the community is tightly clusterized, suppression of rewiring suppresses well-behaved neutral change. (Note that in this figure, in contrast to previous ones, both the σ axis and the μ axis have been inverted to better exhibit the dip in the high-σ regime.)

that speakers stabilize and do not change after initial acquisition.[5] With this assumption, a tightly clusterized network gives rise to a central hub consisting of speakers who are connected to most other speakers in the network, and whose role in the competition of linguistic variants depends on whether the network is rewired or not. With rewiring, in a highly clusterized network new speakers always receive many connections to these central speakers, who, thanks to the critical period assumption, do not themselves change after maturation. Central speakers therefore become the vehicle of change, conserving their own variant while distributing it to speakers newly joined to the network. If network rewiring is suppressed, however, the central speakers of a clusterized network effectively sample from the majority of the population and thus get a very representative picture of the frequency of variants that exist in the network. The central speakers, rather than advancing a change, serve to hinder changes in this setting: as the frequency of an innovation is necessarily low, any innovation event is likely to be quelled by speakers in the central hub, as when these speakers do update their variant, they are unlikely to adopt the innovatory one.

This observation, then, reveals that interactions between features of within-speaker dynamics (here, the critical period assumption) and between-speakers dynamics (here, the degree of clusterization of the social network) may be important enough to affect causation in language change, by adjusting the probability of an innovation surviving and propagating through a language community.

[5] I am much indebted to an anonymous reviewer for raising this point.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

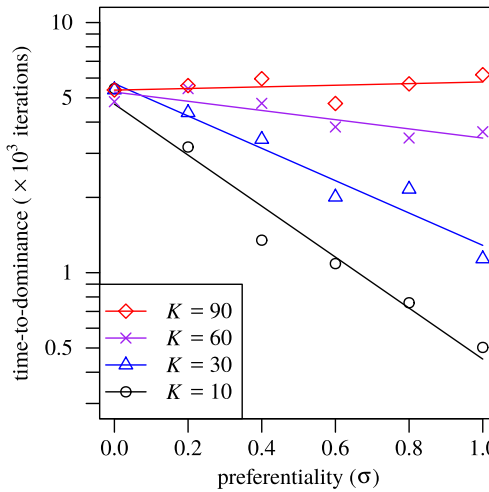

Figure 13
A log–lin plot of time-to-dominance for various values of attachment set size K and preferentiality σ, quantified as the number of iterations it takes for an innovatory variant to permeate the community from an initial state where a number m_0 of speakers entertain the innovatory variant. Here, for each pair of K and σ, the network size was fixed at $N = 100$ and the number of innovators at $m_0 = 10$, and the latter were picked uniformly at random from among all speakers. $C = 3$ competing variants were assumed throughout with innovation rate $\mu = 0.01$. Time-to-dominance is found to be an exponential function of σ, so that increasing σ leads to a speed-up in change for small K.

5.5 *Rate of change*

A comparison of Figures 2 and 3 suggests, impressionistically, that the speed with which an innovation spreads through a community can depend quite drastically on the structure of the community. To investigate this dependence systematically, a final batch of simulations was run (200 simulations for each combination of model parameters), this time with a number of innovative speakers inserted 'by hand' into an otherwise homogeneous community at the start of each simulation. Out of all simulation histories thus generated, the ones where change from this initial state to a state of δ-dominance with $\delta = 0$ by the innovative variant occurred were then selected for further investigation by recording the number of iteration steps it took the community to traverse from the former state to the latter. Figure 13 gives this TIME-TO-DOMINANCE for various combinations of K and σ for a network of size $N = 100$, with 10 innovators. We find that the presence of a central, well-connected hub of speakers in the network has the effect of speeding up change; for small K, the decrease in time-to-dominance is as much as tenfold when moving from $\sigma = 0$ (no clusterization) to $\sigma = 1$ (maximal clusterization).

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

6. Discussion

The above simulations show that the form a linguistic trajectory assumes – whether well-behaved or not – can depend crucially on the social structure and social dynamics of the language community, if none of the competing linguistic variants are biased over others. The results demonstrate well-behaved neutral change for certain types of preferentially attached societies, and show that such change is much less likely in societies lacking preferential connections. Whether real language communities exist with these parameter settings is an empirical matter; the above considerations imply that if such communities exist, well-behaved neutral change is a characteristic property of them.

It is worthwhile to point out explicitly how these results differ from earlier ones, particularly those obtained by Fagyal et al. (2010). While both studies investigate the role of network effects in language change, the model here studied is neutral in the sense that variant acquisition is determined by frequency and does not depend on sociolinguistic considerations. In the model of Fagyal et al. (2010), by contrast, speakers give more weight to speakers who have high degree centrality, so a linguistic variant becomes the fitter the more it is adopted by such central speakers, and their model is thus classified as weighted interactor selection in the Blythe–Croft taxonomy (see Section 2, above). This difference has non-trivial sociolinguistic implications. With a biased model, one assumes that speakers are able to evaluate the centrality or prestige, or both, of each speaker to whom they are connected, and that they in fact pay attention to such evaluations. In a neutral model, the only causative social factor in language change is the way in which speakers are (happen to be) connected, and one need not (or does not) assume that speakers have access to or make use of prestige evaluations.

An important feature of the framework adopted in this paper is not exhibited by previous mathematical models of language change: it models evolution ON and OF a network simultaneously. Infinite-population models have considered non-overlapping, well-mixed generations (e.g. Niyogi & Berwick 1997, Yang 2000, Komarova, Niyogi & Nowak 2001, Mitchener 2006, Niyogi & Berwick 2009), and in most if not all finite-population models (including those of Ke et al. 2008, Fagyal et al. 2010 and Blythe & Croft 2012) the social network is not allowed to evolve as the linguistic variants compete on that network. In the present model, the generations of speakers are overlapping and the network is updated in accordance with the socialization algorithm in use, at each iteration. The simulation results demonstrate that this interplay of the social network rewiring dynamics and the linguistic variant dynamics has an effect on the probability of a language community shifting, as well as on the well-behavedness of any such shifts (Section 5.4); importantly, this refutes previous claims (based on static population modelling) that neutral change cannot be well-behaved (Fagyal et al. 2010, Blythe & Croft 2012).

An obvious criticism of the model is that there is, as yet, no independent evidence for the sort of social network structure the model presupposes. Although the

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

role and importance of social network effects in language change have been noted before (Milroy 1980, Milroy & Milroy 1985), we still lack a deep understanding of the basic properties of human social networks, both topological and dynamic. Two immediate goals can be discerned in this regard. First, empirical studies are needed to establish what the connectivity patterns of actual language communities are – how exactly they are clusterized, what their typical degree distributions are, whether they are possibly multiplex, whether inter-speaker links have weights on them or a binary characterization is sufficient, and so on. Second, these patterns have to be captured in mathematical models that are considerably more complex than the algorithms currently in use in the complex systems and network science literature (for a review of the state of the art and some suggestions for future directions, see Kivelä et al. 2014).

That said, it is possible to interpret the present model, in what is perhaps a promising and productive way, in the light of earlier proposals concerning social factors in linguistic change. We have seen that the preferentiality parameter σ controls the clusterization of the social network, and it is possible to take this as an operationalization of the degree to which a language community is CLOSEKNIT, in the terminology of Milroy & Milroy (1985): networks with large (close to 1) σ will then correspond to communities that are closeknit. Now, we may well imagine several such communities to be connected along inter-community links, composing thereby a network of networks, so that many links are found within the subcommunities, but between the subcommunities a much smaller number of links exist. The intra-community links can then be thought to correspond to the Milroys' STRONG TIES, the inter-community links corresponding to WEAK TIES. In the present model, networks with large σ act as both strong conservers and rapid distributors of linguistic variants: for instance, it can be shown that in the limiting case of $\sigma = 1$, the probability of a speaker in the central cluster of the highly clusterized network distributing their variant to at least one other speaker during the former's lifetime is given by

$$q = 1 - \left(1 - \frac{1}{K} + \mu\left(\frac{1}{K} - \frac{1}{C}\right)\right)^N, \tag{1}$$

as long as $\mu < 1/K$. Importantly, this number is bounded from below by $1 - 1/e \approx 0.63$, irrespective of the values of N (network size) and K (attachment set size), and tends to 1 as K tends to 1 and μ tends to 0. Thus, IT IS ALWAYS MORE PROBABLE FOR VARIANTS FLOWING FROM THE CENTRE OF THE NETWORK TO BE REPLICATED THAN NOT TO BE REPLICATED, and the probability is the greater the more clusterized the network is (Figure 14). This explains both conservatism and progressivism: on the one hand, if no innovatory variants happen to be introduced into the centre of the strongly clusterized network, the centre acts as a strong suppressor against innovations that occur in intermediately (but not strongly) connected speakers, and on the other hand, if an innovatory variant happens to invade the centre of the network, it is almost certainly distributed to

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

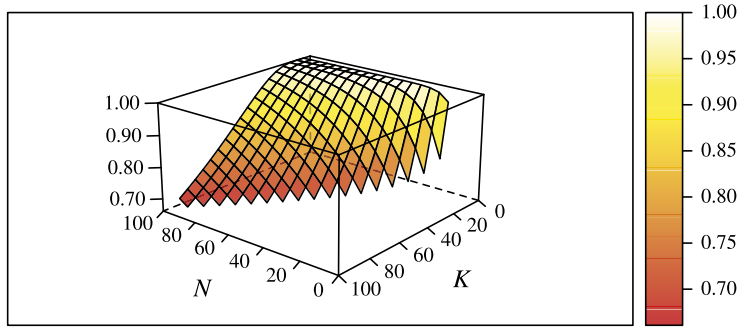

Figure 14
The probability, q, of a central speaker distributing their variant to at least one other speaker before the former is removed from the network by the network-rewiring algorithm, for innovation rate $\mu = 0.01$ and number of competing variants $C = 30$ (Equation (1)). Note that $q \to 1$ as $K \to 1$ and $\mu \to 0$, and that $q > 1 - 1/e \approx 0.63$ for any choice of K and N satisfying $\mu < 1/K$.

at least one other speaker before the bearer of that innovatory variant is removed from the network by the network-rewiring process.

This analogy between the present model and the Milroys' framework can be pressed further. Milroy & Milroy (1985) draw, following Rogers & Shoemaker (1971), a distinction between the INNOVATORS and the EARLY ADOPTERS of a change. In the present model, all innovation events occur in speakers whose degree is K; these speakers, who correspond to the Milroys' innovators, do not belong to the central cluster of the social network. Clearly, language change only happens if, following this initial ACTUATION of an innovatory variant, the variant is subsequently PROPAGATED through the layers of the social network and becomes, eventually, dominant. In the present model, this happens typically if the social network comes to be so rewired that the innovating speaker is 'promoted' to the centre of the clusterized, closeknit community, i.e. if their degree increases due to rewirings of other speakers; this occurs with a finite probability which increases as σ is increased. Once in the centre, the probability of this innovating speaker influencing the variant adoption processes of new speakers is significantly increased; these speakers adopting the new variant then correspond to the Milroys' early adopters, and propagation of the innovatory variant is successful if the number of early adopters is large enough.

Yet the present model does not serve merely as a computational implementation or (partial) corroboration of the Milroys' framework; it adds a positive contribution thanks to the neutrality assumption. As I have noted above (see Section 2), Milroy & Milroy (1985) assume that innovatory variants MUST have a non-zero prestige value attached to them, if they are to propagate successfully through a language community. This is PRIMA FACIE puzzling, as it raises the further question of how (and why) language communities should be able to agree on the social valuation of invading variants.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

> The puzzle is of course how young people living in the closed communities of Ballymacarrett, Clonard and Hammer, whose contact with others outside their areas has been only of a very tenuous kind, have come to reach cross-community consensus on the social value to be assigned to the two variants of the (*pull*) variable (Milroy & Milroy 1985: 374).

The above simulation results suggest that such cross-community consensus may, in fact, be unnecessary. Prestige need not be attached EITHER to linguistic variants OR to individual speakers; in order to have well-behaved neutral change, it suffices to have a non-uniform, but dynamic, population structure containing hubs of speakers.[6] Prestige reduces to degree centrality: the influence of individual speakers lies in the number of connections they have in their language community, not in a social evaluation assigned on top of that number of connections.

7. CONCLUSION

In this paper, I have investigated the possibility that language change is, in some cases, neutral and not motivated by functional, social, articulatory or other biases. I have defined a simple model of variant competition in a finite network of speakers in which variant adoption is neutral, and have tested this model against three criteria that together constitute well-behavedness of change, namely dominance, shifting ability and monotonicity. Results from computer simulations show that if the network of speakers is suitably clusterized, so that it has a central component with some very well connected speakers, well-behaved neutral change is observed in this model. I have proposed a way of interpreting this finding in the framework of Milroy & Milroy (1985) and have suggested that a neutral mechanism, such as the one here considered, calls for a re-evaluation of the role of prestige as a causal factor in at least some cases of change. I have stressed the importance of approaching language diachrony from the viewpoint of mathematical models, and the need to increase the complexity and realism of these models, and hope that results like those reported in this paper can go some way towards justifying this angle of attack. Subsequent work on the neutrality hypothesis should both incorporate more realistic models of social dynamics and relax some of the simplifying assumptions made in this paper, to see whether well-behaved neutral change continues to be observed under such modifications.

The discrepancy seen between the results here reported and those obtained by Ke et al. (2008), Fagyal et al. (2010) and Blythe & Croft (2012) is explained by the different assumptions that go into the definition of each of these four models of language change. In the latter three models, the social network structure underlying the linguistic variant dynamics is not allowed to evolve during a simulation run, so that speakers' neighbourhoods remain fixed. In the present model,

[6] Assuming again, as the model does, that speakers are categorical and invariant after a critical period. The results in Section 5.4 suggest that the interaction of this assumption with the (language-external) social dynamics of the language community is non-trivial; the consequences of relaxing the assumption need to be systematically investigated in future research.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

speakers are added to and removed from the social network in accordance with the network-rewiring algorithm described in Section 3 and Appendix A, and the neighbourhood of a speaker may change during their lifetime if speakers in that neighbourhood are removed, or if new speakers are added thereto. The simulation results show that this interplay between the network-rewiring dynamics and the linguistic variant dynamics, together with the assumption of having speakers who latch onto one or another variant early on and do not change thereafter, is instrumental in supporting well-behaved neutral change in communities that are tightly clusterized (Figure 12).

The model here studied makes a number of predictions which are, in principle, open to investigation and empirical testing. First, the above simulations predict that increasing the number of linguistic variants available to speakers makes neutral change more likely if the innovation rate has a moderate value – but only at the expense of a slight drop in the well-behavedness of change, when quantified using the notions of dominance and monotonicity (Section 5.2). Second, the simulations predict that stable variation should be more likely in clusterized communities than in well-mixing ones (Section 5.3). Finally, change in a clusterized community is much faster than in a well-mixing one of corresponding size (Section 5.5).

The possibility of well-behaved neutral change has implications for diachronic work that seeks to establish non-neutral motivations for language change. While the possibility of neutral change does not imply its probability and does not, PER SE, undermine non-neutral theory in instances where sound reasons exist for believing in the presence of non-neutral motivations, the results here reported do warn against appealing to non-neutral explanations when such reasons are lacking; 'these possibilities [. . .] need to be considered before any claim for 'function' can be made for either variation or change' (Lass 1997: 354). Any particular case of change may in fact be a constellation of neutral and non-neutral factors, and one important goal for research in language diachrony must be to tease apart the relative contributions of these two modes of change.

APPENDIX A. FORMAL DEFINITION OF THE MODEL

Consider a language community of N speakers distributed on an undirected graph (V, E_t), where $V = \{1, \ldots, N\}$ is the set of speakers (vertices) and E_t is an irreflexive, symmetric relation giving the speaker adjacencies (edges), indexed for time t. Denote by $E_t(i) = \{j \in V : (i, j) \in E_t\}$ the NEIGHBOURHOOD of speaker i and by $\deg_t(i) = |E_t(i)|$ the DEGREE of speaker i at time t. Let $\mathcal{C} = \{1, \ldots, C\}$ be the set of linguistic variants, and for each time t define a function $v_t : V \to \mathcal{C}$ which gives the variant of speaker i at time t. Then define an indicator function

$$\chi_t(i, r) = \begin{cases} 1 & \text{if } v_t(i) = r, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

and let the graph (V, E_t) be rewired in discrete time by the following algorithm.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

Algorithm 1. Define a stochastic process to shuffle the graph (V, E_t) as follows.

(1) Let $0 \leqslant \mu, \sigma \leqslant 1$ and K be a positive integer with $K \leqslant N - 1$.
(2) At time 0, the relation E_0 is initialized randomly; say, every speaker has a probability of $1/2$ to be connected to any other speaker.
(3) Choosing a simulation length n, iterate from $t = 1$ to $t = n$.

   (a) Select a speaker i from V uniformly at random.
   (b) Remove all of i's connections.
   (c) For each $d = 0, \ldots, N - 1$, take each speaker other than i having a degree of exactly d; put these speakers into a set Q_d; and shuffle Q_d to make an ordered tuple $\widehat{Q}_d$.
   (d) Define an ordered set Q, the QUEUE, as follows, where $\circ$ denotes concatenation:

$$Q = \widehat{Q}_{N-1} \circ \widehat{Q}_{N-2} \circ \cdots \circ \widehat{Q}_0. \tag{3}$$

   (e) Give the speaker i a connection as follows.

      i. With probability σ, connect i to the first speaker in Q, and delete this speaker from Q
      ii. With probability $1 - \sigma$, connect i to a speaker selected uniformly at random from Q, and delete this speaker from Q.

   (f) Repeat the previous step until i has received exactly K connections.
   (g) Set the variant of speaker i as follows: for each possible variant r, the probability of setting $v_t(i) = r$ is to equal

$$\frac{\mu}{C} + \frac{1 - \mu}{K} \sum_{j \in E_t(i)} \chi_t(j, r). \tag{4}$$

APPENDIX B. QUANTIFYING WELL-BEHAVEDNESS

In quantifying well-behavedness of change, our interest is in how the frequencies of the C competing variants unfold in time. For this, let $x_r(t)$ denote the relative frequency of the rth variant at time t, and let $\vec{x}(t) = (x_1(t), \ldots, x_C(t))$ be the FREQUENCY-STATE of the system. A sequence of frequency-states $\vec{x}(1), \ldots, \vec{x}(n)$ I shall call a HISTORY or (FREQUENCY) TRAJECTORY.

B.1 *Dominance*

Let $0 \leqslant \delta \leqslant 1$. I shall call a frequency-state $\vec{x}(t) = (x_1(t), \ldots, x_C(t))$ δ-DOMINANT if $x_r(t) \geqslant 1 - \delta$ for some r. Dominance times for a history $\vec{x}(1), \ldots, \vec{x}(n)$ are then obtained by the time-averaged measure

$$D_\delta = \frac{1}{n} \sum_{t=1}^{n} \Delta_\delta(t), \tag{5}$$

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

where

$$\Delta_\delta(t) = \begin{cases} 1 & \text{if } \vec{x}(t) \text{ is } \delta\text{-dominant,} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

B.2 *Shifting*

To measure shifting ability, I shall record, for a given simulation run, the number of shifts from δ-dominance by variant r to δ-dominance by another variant $r' \neq r$, for a predefined dominance threshold δ. More formally, for a history $\vec{x}(1), \ldots, \vec{x}(n)$, the shifting measure, S_δ, is defined as the number of time points $t \in \{1, \ldots, n\}$ such that $x_r(t) \geqslant 1 - \delta$ for some t, some r, and $x_{r'}(t') \geqslant 1 - \delta$ for some $t' < t$, some $r' \neq r$.

B.3 *Monotonicity*

A sequence $x(1), \ldots, x(n)$ is monotone if $t < t'$ implies either $x(t) \leqslant x(t')$ or $x(t) \geqslant x(t')$. A history $\vec{x}(1), \ldots, \vec{x}(n)$ will be called monotone if each variant frequency sequence $x_r(1), \ldots, x_r(n)$ is monotone.

Generally, it is possible to estimate the monotonicity of a history by the following measure, for integer $\tau > 0$ and real $\alpha > 0$:

$$W_{\tau,\alpha} = \frac{1}{n-\tau} \sum_{t_0=1}^{n-\tau} \sum_{r=1}^{C} \left(\underbrace{\left(\sum_{t=t_0}^{t_0+\tau-1} s_r^+(t) \right)}_{=m_r^+(t_0,\tau)} \underbrace{\left(\sum_{t=t_0}^{t_0+\tau-1} s_r^-(t) \right)}_{=m_r^-(t_0,\tau)} \right)^\alpha, \tag{7}$$

where

$$s_r^+(t) = \begin{cases} 1 & \text{if } x_r(t) < x_r(t+1), \\ 0 & \text{if } x_r(t) \geqslant x_r(t+1) \end{cases} \tag{8}$$

and

$$s_r^-(t) = \begin{cases} 1 & \text{if } x_r(t) > x_r(t+1), \\ 0 & \text{if } x_r(t) \leqslant x_r(t+1). \end{cases} \tag{9}$$

(For an intuitive characterization of this equation in terms of the quantities $m_r^+(t_0, \tau)$ and $m_r^-(t_0, \tau)$, see Section 4.) This has the following properties under our model.

Proposition 1. *For a simulation operating under Algorithm 1 (Appendix A):*

(i) $0 \leqslant W_{\tau,\alpha} \leqslant \tau^{2\alpha}/2^{2\alpha-1}$ *for all* τ, α;
(ii) $W_{\tau,\alpha} = 0$ *for any* τ *and* α *if and only if the history is monotone;*
(iii) $W_{\tau,\alpha} = 0$ *for some (sufficiently small)* τ *and all* α *if and only if the history is piecewise monotone;*

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

(iv) *$W_{\tau,\alpha} = \tau^{2\alpha}/2^{2\alpha-1}$ for τ even, and $W_{\tau,\alpha} = (\tau^2-1)^{\alpha}/2^{2\alpha-1}$ for τ odd, if and only if the history zig-zags persistently;*

(v) *for large τ, the expected value of $W_{\tau,\alpha}$ is $(2/3)^{2\alpha}\tau^{2\alpha}/C^{2\alpha-1}$ if the history is a random walk.*

Proof. Let $m^+(r, t_0, \tau) = m_r^+(t_0, \tau) = \sum_{t=t_0}^{t_0+\tau-1} s_r^+(t)$ and $m^-(r, t_0, \tau) = m_r^-(t_0, \tau) = \sum_{t=t_0}^{t_0+\tau-1} s_r^-(t)$.

(i) That $W_{\tau,\alpha} \geqslant 0$ is plain. The maximum is achieved when two variants r_1 and r_2 alternate in upward and downward inflections. For τ even, this means that $m^+(r_i, t_0, \tau) = m^-(r_i, t_0, \tau) = \tau/2$ for $i = 1, 2$, for all t_0, and $m^+(r_i, t_0, \tau) = m^-(r_i, t_0, \tau) = 0$ for $i \neq 1, 2$, and therefore

$$W_{\tau,\alpha} = \frac{1}{n-\tau}\sum_{t_0=1}^{n-\tau} 2\left(\left(\frac{\tau}{2}\right)^2\right)^{\alpha} = \frac{\tau^{2\alpha}}{2^{2\alpha-1}}. \tag{10}$$

If τ is odd, then

$$\begin{cases} m^+(r_i, t_0, \tau) &= \dfrac{\tau-1}{2}, \\ m^-(r_i, t_0, \tau) &= \dfrac{\tau-1}{2}+1, \\ m^+(r_j, t_0, \tau) &= \dfrac{\tau-1}{2}+1, \\ m^-(r_j, t_0, \tau) &= \dfrac{\tau-1}{2}, \end{cases} \tag{11}$$

either for $i = 1,\ j = 2$ or for $i = 2,\ j = 1$. In either case,

$$W_{\tau,\alpha} = \frac{1}{n-\tau}\sum_{t_0=1}^{n-\tau} 2\left(\frac{\tau-1}{2}\left(\frac{\tau-1}{2}+1\right)\right)^{\alpha} = \frac{\left(\tau^2-1\right)^{\alpha}}{2^{2\alpha-1}} < \frac{\tau^{2\alpha}}{2^{2\alpha-1}}. \tag{12}$$

(ii) If a history is monotone, then either $m^+(r, t_0, \tau) = 0$ or $m^-(r, t_0, \tau) = 0$ or both for each variant r, for each time t_0. Hence, $W_{\tau,\alpha} = 0$. Conversely, if $W_{\tau,\alpha} = 0$, then $m^+(r, t_0, \tau) = 0$ or $m^-(r, t_0, \tau) = 0$ for each r, for each t_0, which implies that the history is monotone.

(iii) Suppose that a history is monotone when viewed through a window of size τ_0. Then, with the above reasoning, $m^+(r, t_0, \tau) = 0$ or $m^-(r, t_0, \tau) = 0$ in such windows, and consequently we have $W_{\tau_0,\alpha} = 0$ for the average. Conversely, if $W_{\tau_0,\alpha} = 0$, the history is piecewise monotone in windows of size at most τ_0.

(iv) This was shown in (i).

(v) Consider an arbitrary variant r at any time t. Then, $x_r(t)$ can inflect upwards in two ways: either r is selected to change so that x_r increases while some $r' \neq r$ decreases, or some $r' \neq r$ is selected so that $x_{r'}$ decreases and x_r increases. Since $\vec{x}(t)$ is assumed to be a random walk,

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

the probability for $x_r(t)$ to increase is then given by

$$p = \frac{1}{C} \cdot \frac{1}{3} + \frac{C-1}{C} \cdot \frac{1}{3} \cdot \frac{1}{C-1} = \frac{2}{3C} \tag{13}$$

(probability of picking r times probability of x_r increasing (rather than decreasing or not changing), plus probability of picking r' times probability of $x_{r'}$ decreasing so that x_r increases). By symmetry, the probability for r to inflect downward is the same. In a window of length τ, we would then expect $p\tau$ upward and $p\tau$ downward inflections for variant r, if τ is sufficiently large. This gives us

$$F := \sum_{r=1}^{C} \left(\left(\frac{2}{3C} \tau \right)^2 \right)^{\alpha} = C \frac{2^{2\alpha} \tau^{2\alpha}}{3^{2\alpha} C^{2\alpha}} = \left(\frac{2}{3} \right)^{2\alpha} \frac{\tau^{2\alpha}}{C^{2\alpha-1}}, \tag{14}$$

hence

$$W_{\tau,\alpha} = \frac{1}{n-\tau} \sum_{t_0=1}^{n-\tau} F = \left(\frac{2}{3} \right)^{2\alpha} \frac{\tau^{2\alpha}}{C^{2\alpha-1}}, \tag{15}$$

as wished. □

Now let

$$M_\tau = 1 - \frac{W_{\tau,1/2}}{\tau}. \tag{16}$$

Then we have the following.

Proposition 2. *For a simulation operating under Algorithm 1 (Appendix A):*

(i) $0 \leqslant M_\tau \leqslant 1$ *for any* τ*;*
(ii) $M_\tau = 1$ *for any* τ *if and only if the history is monotone;*
(iii) $M_\tau = 1$ *for some (sufficiently small)* τ *if and only if the history is piecewise monotone;*
(iv) $M_\tau = 0$ *for even* τ *if and only if the history zig-zags persistently;*
(v) *for large* τ*, the expected value of* M_τ *is* $1/3$ *if the history is a random walk.*

Proof. From Proposition 1 by simple substitution via (16). □

Thus, the value of M_τ will range from 0 (inclusive) to 1 (inclusive) for even window sizes τ. The closer this value is to 1, the more monotone the history is; the closer this value is to 0, the less monotone the history is. Having these desirable properties, M_τ (restricted, without loss of generality, to even τ) will serve as our measure of monotonicity.

REFERENCES

Alonso, David, Rampal S. Etienne & Alan J. McKane. 2006. The merits of neutral theory. *Trends in Ecology and Evolution* 21.8, 451–457.
Anttila, Raimo. 1989. *Historical and comparative linguistics*, 2nd edn. Amsterdam: Benjamins.

https://doi.org/10.1017/S0022226716000141 Published online by Cambridge University Press

Bailey, Charles-James N. 1973. *Variation and linguistic theory*. Arlington, VA: Center for Applied Linguistics.
Barabási, Albert-László & Réka Albert. 1999. Emergence of scaling in random networks. *Science* 286, 509–512.
Baxter, Gareth J., Richard A. Blythe, William Croft & Alan J. McKane. 2006. Utterance selection model of language change. *Physical Review* E 73, 046118.
Baxter, Gareth J., Richard A. Blythe, William Croft & Alan J. McKane. 2009. Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21.2, 257–296.
Blythe, Richard A. & William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88.2, 269–304.
Coussé, Evie & Gert De Sutter. 2012. De historische wortels van de rode en groene volgorde in het Nederlands. *Taal en Tongval* 64, 73–101.
Croft, William. 2000. *Explaining language change: An evolutionary approach*. Harlow: Longman.
Denison, David. 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for language change*, 54–70. Cambridge: Cambridge University Press.
Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser & Kiran Lakkaraju. 2010. Centers and peripheries: Network roles in language change. *Lingua* 120, 2061–2079.
Ghanbarnejad, Fakhteh, Martin Gerlach, José M. Miotto & Eduardo G. Altmann. 2014. Extracting information from S-curves of language change. *Journal of the Royal Society Interface* 11, 20141044.
Gross, Thilo, Carlos J. Dommar D'Lima & Bernd Blasius. 2006. Epidemic dynamics on an adaptive network. *Physical Review Letters* 96, 208701.
Harrington, Jonathan. 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34, 439–457.
Hawkins, John A. 1990. Seeking motives for change in typological variation. In William Croft, Keith Denning & Suzanne Kemmer (eds.), *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday*, 95–128. Amsterdam: Benjamins.
Hubbell, Stephen B. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton, NJ: Princeton University Press.
Hull, David L. 1988. *Science as a process*. Chicago, IL: University of Chicago Press.
Itkonen, Esa. 1981. Rationality as an explanatory principle in linguistics. In Horst Geckeler, Brigitte Schlieben-Lange, Jürgen Trabant & Harald Weydt (eds.), *Logos semantikos: Studia linguistica in honorem Eugenio Coseriu 1921–1981,* vol. 2, 77–87. Berlin: De Gruyter.
Itkonen, Esa. 1982. Short-term and long-term teleology in linguistic change. In J. Peter Maher, Allan R. Bomhard & E. F. Konrad Koerner (eds.), *Papers from the 3rd International Conference on Historical Linguistics*, 85–118. Amsterdam: Benjamins.
Ke, Jinyun, Tao Gong & William S.-Y. Wang. 2008. Language change and social networks. *Communications in Computational Physics* 3.4, 935–949.
Kerswill, Paul. 1996. Children, adolescents, and language change. *Language Variation and Change* 8, 177–202.
Kimura, Motoo. 1994. *Population genetics, molecular evolution, and the neutral theory*. Chicago, IL: The University of Chicago Press.
Kivelä, Mikko, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno & Mason A. Porter. 2014. Multilayer networks. *Journal of Complex Networks* 2, 203–271.
Komarova, Natalia L., Partha Niyogi & Martin A. Nowak. 2001. The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology* 209, 43–59.
Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1.3, 199–244.
Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.
Lightfoot, David. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.
Lightfoot, David. 1999. *The development of language: Acquisition, change, and evolution*. Malden, MA: Blackwell.
Milroy, James & Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21.2, 339–384.
Milroy, Lesley. 1980. *Language and social networks*. Oxford: Blackwell.

Mitchener, W. Garrett. 2006. A mathematical model of the loss of verb-second in Middle English. In N. Ritt, H. Schendl, C. Dalton-Puffer & D. Kastovsky (eds.), *Medieval English and its heritage*, 189–202. Frankfurt am Main: Peter Lang.
Nahkola, Kari & Marja Saanilahti. 2004. Mapping language changes in real time: A panel study on Finnish. *Language Variation and Change* 16, 75–92.
Niyogi, Partha & Robert C. Berwick. 1997. A dynamical systems model for language change. *Complex Systems* 11, 161–204.
Niyogi, Partha & Robert C. Berwick. 2009. The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences* 106.25, 10124–10129.
Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (eds.), *Language change: Contributions to the study of its causes*, 173–198. Berlin: Mouton de Gruyter.
Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan L. Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam: Benjamins.
Postal, Paul M. 1968. *Aspects of phonological theory*. New York, NY: Harper & Row.
Postma, Gertjan. 2010. The impact of failed changes. In Anne Breitbarth, Christopher Lucas, Sheila Watts & David Willis (eds.), *Continuity and change in grammar*, 269–302. Amsterdam: Benjamins.
Reeve, H. Kern & Laurent Keller. 1999. Levels of selection: Burying the units-of-selection debate and unearthing the crucial new issues. In Laurent Keller (ed.), *Levels of selection in evolution*, 3–14. Princeton, NJ: Princeton University Press.
Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
Rogers, Everett M. & F. Floyd Shoemaker. 1971. *Communication of innovations*, 2nd edn. New York, NY: Free Press.
Sankoff, Gillian & Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83.3, 560–588.
Traulsen, Arne, Francisco C. Santos & Jorge M. Pacheco. 2009. Evolutionary games in self-organizing populations. In T. Gross & H. Sayama (eds.), *Adaptive networks: Theory, models and applications*, 253–267. Cambridge, MA: NECSI.
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, 241–254.
Vennemann, Theo. 1993. Language change as language improvement. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives*, 319–344. London: Longman.
Wallenberg, Joel C. 2013. A unified theory of stable variation, syntactic optionality, and syntactic change. Talk delivered at the 15th Diachronic Generative Syntax (DiGS) Conference, University of Ottawa, August 2, 2013.
Yang, Charles D. 2000. Internal and external forces in language change. *Language Variation and Change* 12, 231–250.

Author's address: *The University of Manchester, Linguistics and English Language, Samuel Alexander Building, Oxford Road, Manchester, M13 9PL, UK*
henri.kauhanen@manchester.ac.uk