Momentum in language change

Citation for published version:
Stadler, K, Blythe, R, Smith, K & Kirby, S 2016, 'Momentum in language change: A model of self-actuating s-shaped curves' Language Dynamics and Change, vol. 6, no. 2, pp. 171-198. DOI: 10.1163/22105832-00602005

Digital Object Identifier (DOI):
10.1163/22105832-00602005

Published In:
Language Dynamics and Change

Momentum in language change: a model of self-actuating s-shaped curves

Abstract

Like other socially transmitted traits, human languages undergo cultural evolution. While humans can replicate linguistic conventions to a high degree of fidelity, sometimes established conventions get replaced by new variants, with the gradual replacement following the trajectory of an *s-shaped curve*. Although modelling work has shown that only a bias favouring the replication of the new linguistic variant can reliably reproduce the dynamics observed in language change, the source of this bias is still debated. In this paper we compare previous accounts with a *momentum-based selection* account of language change, a replicator-neutral model where the popularity of a variant is modulated by its *momentum*, i.e. its *change in frequency of use* in the recent past. We present results from a multi-agent model that are characteristic of language change, in particular by exhibiting spontaneously generated s-shaped transitions. We discuss several empirical questions raised by our model, pertaining to both momentum-based selection as well as previous accounts of language change.

keywords: language change; cultural evolution; momentum; age vectors; s-shaped curves

1 Introduction

Human languages are a prime example of culturally evolving traits: they are made up of socially learned conventions which are constantly being replicated and which exhibit great diversity across the globe (Evans and Levinson, 2009). Important aspects of the dynamics of language change are well-understood. Firstly, language change is *sporadic* (de Saussure, 1959; Labov, 2001). Of all the conventions that make up a single language, at any given point most of them are not undergoing change, but are replicated faithfully, from basic word order patterns down to the pronounciation details of individual words (Pierrehumbert, 2002). Languages are transmitted robustly over many generations, a necessary requirement for their use as a tool for communication (Lewis and Laland, 2012). Secondly, when a convention *does* change, individuals will gradually replace an established variant with a new variant. This gradual replacement exhibits directed transitions in the form of *s-shaped curves* such as in Fig. 1, akin to the patterns of logistic growth found in biological evolution (Bailey, 1973; Altmann et al., 1983; Kroch, 1989; Denison, 2003; Blythe and Croft, 2012)[1]. This similarity to the signature of adaptive selection in biology is puzzling: linguistic conventions are *arbitrary*, which means we should not expect an inherent advantage in particular linguistic variants, such as which basic word order is used by a language, or how exactly a distinctive phonemic segment is pronounced (as long as it maintains its contrastive function). How and why would an entire population of speakers go about replacing an existing convention with a different one "to say the same thing"?

[Figure 1 about here.]

1.1 Language-internal accounts

In order to explain *why* languages change, many studies have attempted to pin down the causes of individual changes by systematically comparing the states of the languages prior to and after a change (Hockett, 1965; McMahon, 1994). While many of the earliest such studies would attribute change to the gradual accumulation of performance and transmission errors alone (Jespersen, 1922; Hockett, 1958), the generativist paradigm with

[1] While the notion of 's-shaped curves' is notoriously ill-defined, for the purposes of this paper it will suffice to use Blythe and Croft's definition as any directed trajectory that does *not* exhibit "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" (2012, p.285).

its focus on the language acquisition device shifted the attention firmly to child-based language change. Studies of language change in the generative tradition have traced changes back to the re-ordering or simplification of rules (Kiparsky, 1968; Wang, 1969; Bailey, 1973; Lass, 1980; Vennemann, 1983), often based on children's erroneous reanalysis of linguistic parameters based on limited linguistic input (e.g. Ellegard (1953); Lightfoot (1979); Kroch (1989); Lightfoot (1991); see Foulkes and Vihman (2013) for a review). Rather than characterising change as the result of imperfect transmission, a more recent strand of research sees language as a *complex adaptive system* which evolves to fulfill the communicative needs of its speakers, while at the same time adapting to the constraints imposed by their learning mechanisms (Kirby, 1999; Steels, 2000; Griffiths and Kalish, 2007; Beckner et al., 2009).

What unites these *language-internal* accounts is that they all rely on a qualitative difference between the language states prior to and after the change. This difference can be based on a variety of factors, such as the languages' expressivity, processing efficiency, or simply their stability with respect to error-prone language acquisition. Within historical and variationist linguistics such explanations of language change have long been criticised on the basis that they *overpredict* change (de Saussure, 1959; Greenberg, 1959; Weinreich et al., 1968; Lass, 1980; Ohala, 1989; Croft, 2000; Labov, 2001; Winter-Froemel, 2008). In their seminal paper, Weinreich et al. succinctly summarised the issue and coined it the *actuation problem*: "Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times?" (Weinreich et al., 1968, p.102).

In other words, language-internal pressures by themselves do not account for the *sporadicity* of language change: many non-adaptive or sub-optimal structures that are claimed to have been selected against in one language will happily persist in other languages – and when they finally do change, language-internal accounts often offer no explanation of what triggered the *actuation* of the change (de Saussure, 1959; Postal, 1968; Ohala, 1993). While language-internal factors offer insights into *what* changes are more likely to occur than others (Jaeger and Tily, 2010; Wedel et al., 2013), they do not explain *when* or *why* the stable transmission of language should suddenly cave under functional pressures. To account for the sporadic nature of language change, many have argued that it is not enough to rely on intra-linguistic factors alone.

1.2 Social accounts

Sociolinguistic research of the past five decades has shown that innovations do not spread uniformly across a given speech community, but that the progression of change is stratified based on factors such as a speaker's age, ethnicity, or socio-economic status (Foulkes and Docherty, 2006; Tagliamonte, 2012). *Social accounts* hold that the social aspects of linguistic variants, rather than their inherent linguistic character, are responsible for driving language change (Sturtevant, 1947; Croft, 2000; Labov, 2001; Croft, 2006). Social accounts of language change are *evolutionary* in nature: they decouple the generation of *variation* from the process of *selection* which leads to the diffusion of variants through a speech community. The underlying mechanisms, however, are very different from biological evolution: while the generation of new variants is assumed to be driven by linguistic or functional factors, social accounts attribute the ultimate *selection* of variants to extra-linguistic social factors (Ohala, 1989; Croft, 2000; Stevens and Harrington, 2013). The 'division of labour' between language-internal and social pressures in this approach can simultaneously account for the arbitrary adoption of one linguistic convention from the pool of variants over another, while at the same time explaining the crosslinguistic distribution of linguistic features which reflect functional pressures.

Recent work on a mathematical model of language change showed that only the presence of a bias favouring the replication of the incoming variant can reliably reproduce the s-shaped transitions observed in language change (Blythe and Croft, 2012). While this mechanism, known as *replicator selection*, is in principle also compatible with language-internal biases, the authors eschew this conclusion. In line with social accounts of language change they conclude instead that it is the *social prestige* of a new variant that is responsible for its preferential replication. Importantly, the sociolinguistic use of the term *prestige* actually refers to a *content bias*: rather than preferentially copying variants used by prestigious individuals, *prestige* is simply another name for a bias that, while social in origin, is actually inherent to the linguistic variant (Sturtevant, 1947; Labov, 2001). Crucially, social accounts do not solve the underlying logical problem of how a population would agree on the selection of a new variant if there is no objective advantage to that variant. The choice of the population to attach preferential prestige to some variant is as arbitrary and requires just as much explanation as a population's increased use of one linguistic variant over another. Because variant prestige is not accounted for within the theory (Meillet, 1921; Labov, 2001) and can only be attributed post-hoc (Sankoff, 1988; Trudgill, 2004), social accounts also make no predictions whether particular changes are likely to happen or not. If we saw competing variants as completely identical in terms of both their linguistic *and* social value, how could directed transitions come about? To address this question, it is useful to consider ideas from the wider domain of cultural evolution.

1.3 Replicator-neutral accounts

The evolutionary approach that has been adopted in the quantitative study of language variation and change is also used widely to study processes of cultural change more generally (Boyd and Richerson, 1985; Mesoudi, 2011). Interestingly, even though replicator-neutral accounts – where individuals have no inherent preference for any of the competing variants – have been studied extensively in the context of cultural evolution (Bentley et al., 2004, 2007), such models have received relatively little attention in the study of linguistic change (e.g. Trudgill, 2008; Baxter et al., 2009).

Among the few attempts to build a bridge between general models of cultural evolution and the dynamics of language change is Reali and Griffiths (2010). Starting from a model of pure neutral evolution by random copying – where individuals replicate the different variants proportionally to their current prevalence – they augment it with a pressure for *regularisation*, i.e. a slight preference for individuals to adopt grammars exhibiting no variation. The authors show that the trajectories produced by this regularising neutral model exhibit s-shaped growth, as long as only those trajectories which start at 0% use of a novel variant and end at 100% use are considered. Crucially, however, their mathematical model captures all possible trajectories between those two points, and their result holds only for the *average of all possible trajectories.* This idealised trajectory is highly unlike the 'typical' transitions produced by neutral evolution, which are characterised by a noisy trajectory with many reversals. The strict symmetry of their Markov model also predicts that for every completed language change we should find an equal amount of actuated changes that went to the 50% mark before being interrupted, a situation does not seem to be the case for language change. These considerations call into question whether neutral evolution by random copying can provide an adequate model of the dynamics of language change (Blythe, 2012).

While in pure neutral evolution models the likelihood of replicating a variant is assumed to be dependent on that variant's current prevalence alone, another class of replicator-neutral models that has received increased attention recently considers the effects of *temporal information* and *memory* on the diffusion of cultural (and particularly linguistic) traits. Labov

(2001) suggested that the systematic incrementation of sound changes across generations could be explained by the notion of *age vectors*. He hypothesises that, following an initial stage where learners acquire the average community usage of linguistic variants, adolescents advance their productions in line with the age stratification of variable usage that can be observed in the population – in other words, it presumes that youngsters have a bias against sounding *outdated*. This idea was taken up by Mitchener (2011), who framed it in terms of *prediction-driven instability*: in his mathematical model, individuals are able to observe the usage levels of a categorical sociolinguistic variable among the 'older' and 'younger' individuals in the population. New individuals entering the population then adopt a usage rate according to the predicted future use of the variants, by extrapolating from the usage levels of the two groups along an idealised logistic curve. While the model exhibits spontaneous transitions between the two (or more) competing language states, it produces trajectories that exhibit rapid growth from the onset of the change, unlike the gradual uptake observed in empirical data such as in Fig. 1. The model also relies on individuals not changing their usage frequencies once they are added to population, i.e. the individuals' usage rates remain completely fixed after they are initally acquired. This leaves open the question of whether the same mechanism could also give rise to directed changes when individuals adjust their usage rates throughout their lifetime, as has been observed in linguistic changes (Sankoff and Blondeau, 2007).

Another general model of cultural evolution based on a similar principle is Gureckis and Goldstone's model of momentum-based selection (Gureckis and Goldstone, 2009), which we will study more closely in the remainder of this paper. In this model, an individual's choice of cultural variants is influenced by the variants' *momentum*, i.e. by *changes to the variants' frequency of use* in the recent past. Individuals are assumed to be biased towards variants which have recently seen an increase in their usage rate, and conversely biased against variants that have been adopted relatively less frequently in the recent past.

They test their model on a dataset of the frequency of names given to children in the US over 127 years. Their prediction for the popularity of a name in a given year, which is based on its long-term popularity modulated by its momentum, leads to a better fit of the empirical data than the prediction made by pure random copying accounts, which is based on its popularity in the previous year alone. Importantly, Gureckis and Goldstone's model was intended to improve the fit of an empirical prediction, but not meant as a generative model of individual behaviour. The authors rule this out, noting that "if rising names are preferred, which in turn causes them to rise, then a momentum bias might quickly lead to convergence on a single token" (Gureckis and Goldstone, 2009, p.668). They regard this as a negative property of the model, as they are interested in mechanisms that exhibit *cycles* in the popularity of traits, such as found in the realm of fashion (Kroeber, 1919; Berger and Le Mens, 2009; Acerbi et al., 2012). In language change, on the other hand, convergence on a single convention is the rule rather than the exception, suggesting that momentum-based selection may be more appropriate as a model for language than for other cultural domains such as first names.

2 Momentum-based selection

Our main contribution in this work is to investigate the dynamics of momentum-based selection by integrating it into an existing framework of language change, and evaluating it with respect to the characteristics of language change we identified above: the sporadic nature of changes which, once actuated, proceed in an orderly, directed manner. We begin by reviewing the original formulation of momentum-based selection in Gureckis and Goldstone (2009). The model is built around tracking exponentially weighted moving averages (EWMAs) of the relative frequencies of competing cultural traits in an unstructured population. Given a time series of relative frequencies $\vec{n} = \langle n_1, n_2, n_3, \ldots \rangle$, the weight of each datapoint towards the moving average, which we denote $\hat{n}_\alpha(t)$, decreases exponentially over time (hence the name). Given a new datum n_t, the moving average can be updated iteratively using

$$\hat{n}_\alpha(t) = \alpha \cdot n_t + (1-\alpha) \cdot \hat{n}_\alpha(t-1) \qquad (1)$$

where the subscript $\alpha \in [0, 1]$ specifies a constant smoothing coefficient that determines the weight given to newly incorporated datapoints, as well as how quickly the datapoints' weight decreases over time. At time t, the relative weight of datum n_{t-i} in the current average is $\alpha \cdot (1-\alpha)^i$. The higher α, the more weight is given to more recent datapoints. Based on this, the momentum of a variant at time t, $m(t)$, is determined by calculating two EWMAs $\hat{n}_\alpha(t), \hat{n}_\gamma(t)$ of the variant's attested frequencyies $\langle n_1 \cdots n_t \rangle$ with decay parameters $\gamma > \alpha$, and taking their difference,

$$m(t) = \hat{n}_\gamma(t) - \hat{n}_\alpha(t). \qquad (2)$$

Because the higher γ gives more weight to recent datapoints, the moving average $\hat{n}_\gamma(t)$ corresponds to the recent popularity of a trait while $\hat{n}_\alpha(t)$

captures its long-running popularity. The momentum term $m(t)$ will consequently be positive if a variant has been more popular in the recent past compared to its long-term popularity, and negative if the variant has been adopted relatively less frequently in the recent past.

2.1 Mathematical properties of the momentum dynamics

To understand just what is captured by the momentum term $m(t)$, we can investigate the general dynamics of the difference between two EWMAs $\hat{n}_\alpha(t), \hat{n}_\gamma(t)$ based on their decay parameters $\gamma > \alpha$. The strongest possible trend in changes to relative variant frequency can be achieved by initialising both EWMAs at one extreme values (e.g. 0), then continuously updating them with the opposite extreme value (i.e. 1). Starting from an initial momentum of zero, both the number of data points it takes to reach the maximum difference between the two and the amplitude of this highest possible momentum value depend on both decay parameters α and γ, as can be seen in Fig. 2a. What is of interest to us are the different shapes of these momentum curves, and how they affect the model dynamics: a parameter combination which exhibits a rapidly rising curve will cause an individual to posit a trend based on just a few suggestive input data points, while a curve that slopes off slowly means that a momentum bias will persist for a long time after the initial detection of a trend. The parameter γ is of particular importance, as it controls the time depth at which trends are detected, as can be seen in Fig. 2b. A high γ causes the momentum term to immediately reflect short-term variation in the input, while settings of γ closer to α lead to more conservative trend estimates which smooth over the noise present in individual input data points. Generally, the number of iterations that both EWMAs have to be updated with the same constant input value before the maximum possible difference between the two is reached is

$$t_{\mathrm{mmax}}(\alpha, \gamma) = \frac{\ln \frac{\alpha}{\gamma}}{\alpha - \gamma}. \tag{3}$$

[Figure 2 about here.]

The maximum possible amplitude of the momentum term at that point is

$$m_{\max}(\alpha, \gamma) = e^{-\gamma t_{\mathrm{mmax}}(\alpha,\gamma)} - e^{-\alpha t_{\mathrm{mmax}}(\alpha,\gamma)} \,. \tag{4}$$

Knowing the mathematical boundaries of the momentum-based selection bias we can now go on to incorporate the momentum bias into a generative model of language change.

2.2 The Utterance Selection Model of language change

To investigate the dynamics of momentum-based selection as a model of individual behaviour, we implemented the momentum-based selection bias in the *utterance selection model* of language change (USM) (Baxter et al., 2006; Blythe and Croft, 2012). Derived from Croft's evolutionary theory of language change (2000), the USM provides a well-studied multi-agent framework to study the dynamics of the competition and diffusion of *discrete* linguistic replicators, be they lexical items, constructions, or different categorical variants of a speech sound[2].

Two fundamental principles underlie the design of the USM: firstly, the individual agents use the competing variants *proportionally*, rather than categorically. In the minimal case with only two competing variants studied here, an agent's usage rates can be fully described by a single number, call it x, in the range $[0, 1]$. While this value can be interpreted as reflecting some cognitive state of the speaker, it also has a more direct behavioural correspondent: when an agent is selected to participate in an interaction, their probability of producing the novel variant is equal to x, while the probability of producing the competing variant is $1 - x$. This aspect of the USM is in line with linguistic evidence which shows that human language use is inherently variable (Kroch, 1994; Labov, 1994; Bybee, 2007).

Secondly, to mimic humans' tendency to *align* their linguistic behaviour with that of their interlocutors, agents continuously tune their own proportion of variant usage towards the productions they observe in interactions with other agents (Jaeger and Snider, 2013; Nardy et al., 2013). This aspect of the USM is in line with the finding that many aspects of linguistic behaviour do not remain fixed throughout an individual's lifetime, instead remaining malleable across the life span (Kerswill, 1996; Sankoff and Blondeau, 2007; Beckner et al., 2009; Bowie and Yaeger-Dror, 2013; Stanford, 2014). According to the formal definition of the USM (Baxter et al., 2006), an agent's current proportion of use of a variant $x_\alpha(t)$, is simply an exponentially weighted moving average (EWMA) of the frequencies of the incoming variant that the agent has observed in their input over time[3]. The

[2]For an account of how age vectors can drive change in a continuous dimension such as vowel productions, see Swarup and McCarthy (2012).

[3]For simplicity of notation we will henceforth omit the ˆ above the variables denoting

rate of alignment is controlled by the decay parameter α of this EWMA, which can be understood as a *learning rate*. This learning rate is typically held small (in the range of 0.01): there is alignment, but the individual frequency adjustments after an interaction are very small and it takes many interactions for an agent to change their preferred variant.

On top of this basic update rule, a USM agent's alignment behaviour can be altered by applying biases to their input data before it gets incorporated into the EWMA. This is where momentum-based selection comes into play.

2.3 Momentum-based selection in the USM

We now explain how to minimally incorporate momentum-based selection as defined by Gureckis and Goldstone (2009) into the USM. Assuming an agent using learning rate α has just engaged in its t-th interaction and observed another agent use the incoming variant with a relative frequency of y, then their own frequency of use x_α is updated to be

$$x_\alpha(t) = \alpha \cdot f(y) + (1-\alpha) \cdot x_\alpha(t-1) \ , \tag{5}$$

where $f(y)$ is a function from $[0,1]$ to $[0,1]$ which transforms the *objective* observed frequency of the variant into a *perceived frequency* which the agent then aligns to. Similar to Gureckis and Goldstone (2009) we can now simply define the perceived frequency $f(y)$ of an agent in the momentum-based USM as the objective frequency y of a variant observed in an interaction offset by that variant's momentum,

$$f(y) = y + b \cdot m'(t) \tag{6}$$

with the exception of

$$f(0) = 0 \quad \text{and} \quad f(1) = 1 \ . \tag{7}$$

We impose the latter since we are only interested in modelling the diffusion of existing linguistic variants, not in how those variants were introduced into the population to begin with – in other words, this constraint stops our momentum-biased selection function from generating novel, unattested variants (Boyd and Richerson, 1985). The positive bias parameter b in equation 6 controls the strength with which the normalised momentum term $m'(t)$ as defined below in equation 8 influences the perceived frequency. Should the momentum bias cause $f(y)$ go below 0 or above 1, it

EWMAs.

is simply truncated at 0 and 1, respectively[4]. Crucially, because the momentum term can be positive or negative (depending on the direction of the trend), this perceived frequency function is *symmetric*, which makes it *replicator-neutral*: no matter what value is used for parameter b, the function does not a priori favour one of the variants over the other.

Since the effect of different strengths of this bias parameter on the model dynamics is relevant to our analysis, we have to make sure that its settings are comparable across conditions. This isn't as straightforward as it might seem, because the range of values that the momentum term $m(t)$ as defined in equation 2 can take on depends on their decay parameters α and γ, as can be seen from Fig. 2. The absolute amplitude of the momentum curves is of little interest to us; on the contrary, the differences in maximum possible amplitude distort the effect of the bias parameter b which is supposed to control the strength with which momentum is applied. To counteract this, we normalise the momentum term $m(t)$ based on the α, γ used in a given simulation condition. For any given pair of decay rates α, γ, we can scale the momentum term to the $[-1, 1]$ range by defining the normalised momentum

$$m'(t) = \frac{x_\gamma(t) - x_\alpha(t)}{m_{\max}(\alpha, \gamma)} \, . \tag{8}$$

To calculate the momentum component in the numerator, the difference between two EWMAs, we simply re-use the agent's own usage frequency, which according to the USM definition is also an EWMA. To augment the basic USM with momentum-based selection, every agent simply has to keep track of another x_γ on top of the long-term estimate x_α it already maintains.

3 Results

3.1 Analytical approximation

Before proceeding to a full population-based simulation we can establish the general dynamics of the model by investigating the behaviour of an individual agent set in a production-perception loop (Wedel, 2006). We initialise a single agent to use the incoming variant at some low level and

[4]The exact form of the bias function $f(x)$ matters much less than its monotonicity and the fact that $f(x) > x$ when the momentum term is positive (i.e. when the agent perceives an upward trend) and $f(x) < x$ when it is negative (indicating a downward trend).

repeatedly update their two EWMAs $x_\alpha(t), x_\gamma(t)$ by having them align to their internal proportion of use $x_\alpha(t)$. Nothing happens: an agent aligning to their own variable use with no added noise simply remains at that proportion, and the momentum term remains 0 (see the first 100 interactions in Fig. 3). To test how the model reacts to fluctuations in the input we alter the agent's input by fabricating a datapoint which suggests that their interlocutors are actually categorically using the incoming variant (see Fig. 3a). When the agent aligns to this usage rate it leads to a small punctual increase in their variant use, but the sudden change in the input data also makes the momentum term take on a positive value (dashed grey line). Following the fabricated data point, the agent again receives their own samples as input data. But the bias exerted by the momentum term, which makes the agent's *perceived* usage rate higher than their actual usage rate, causes further increases in their use of the incoming variant. However, the lack of further perturbations causes the momentum to decay back towards 0, and the agent becomes stationary again at a usage level not far from their initial setting. If we introduce a second fabricated data point shortly after the first one, the model's behaviour changes dramatically: the system enters a regime where the momentum bias generated by the two fabricated datapoints affects the perceived frequency of the agent's input so much that it causes the momentum term to increase further, leading to self-reinforcing runaway change (Fig. 3b).

[Figure 3 about here.]

This preliminary analysis shows that the momentum-based selection model exhibits two different regimes, accounting for both periods of stability and of directed change. Capturing the dynamics of the transition between the two regimes is however not trivial: particularly the switch from a period of stability to a directed transition depends crucially on both the strength of the momentum bias as well as random fluctuations in the agents' input as they sample input data from their interlocutors. We therefore turn to numerical simulations, where the data production and agent interactions will be driven by stochastic processes.

3.2 Numerical simulation

In order to get a fuller picture of the momentum-based selection dynamics we explored a performed simulations with a total of $2,520$ parameter combinations. The six parameters of the momentum-based USM are summarised below. Only one, the learning rate α, was held constant across all simulation runs, the other five parameters were varied at the levels given in parentheses:

- α: the agents' learning rate (.01)
- γ: the agents' short-term memory decay rate $(.015, .02, .025, .03, .35, .4)$
- T: the Binomial sample size determining the resolution at which agents can observe each other's relative usage frequencies $(2, 3, 4, 5)$
- b: the bias strength with which agents apply the normalised momentum to yield their *perceived* frequency of usage $(.5, 1.0, 1.5, 2.0, 2.5)$
- N: number of agents in the population $(2, 5, 10, 20, 30, 50, 100)$
- x_0: initial proportion of the incoming variant used by all agents $(.01, .02, .03)$

Combining all these possible parameter combinations and running the $2,520$ conditions for 48 trials each resulted in a total of $120,960$ simulation runs. On top of the conditions listed above, we also produced simulation runs where we set the bias strength $b = 0$, which makes it equivalent to pure neutral evolution. $24,192$ runs from this additional condition provide a baseline that the dynamics of our momentum-based selection model can be compared against. Every run of our simulations proceeds as follows:

Firstly, initialise N agents, setting both their $x_\alpha(0)$ and $x_\gamma(0)$ to x_0. Then, carry out interactions between agents by repeating the following steps:

1. randomly select two agents i, j from the pool of N agents – we assume that all pairs of agents have the same probability of interacting with each other.
2. let both agents produce T tokens of the variable by taking a random sample n_i, n_j for each agent from the Binomial distribution $B(T, x_\alpha)$, using the agents' respective value of x_α at the time of the interaction.
3. calculate the perceived frequencies that the agents will align to, using equation 6. For agent i, who will align to j's productions, calculate $f(\frac{n_j}{T})$ using agent i's current normalised momentum term $m'(t)$; for agent j, calculate $f(\frac{n_i}{T})$ using j's $m'(t)$.
4. update both agents' x_α as well as x_γ by incorporating their perceived frequency according to equation 5.

The simulations were run until every individual in the population had converged to within one millionth of a percent of using only one of the two competing variants, or for a maximum of $200,000$ interactions per agent[5].

3.3 Simulation results

For the sake of our analysis we use a simple definition of what a 'transition' is. Taking a fixed threshold (say 5%), we can define the two extreme areas where the mean population usage level of the minority variant is below this threshold as the two regions of 'near-categorical use' of either variant. A transition, then, is the period in which the mean usage levels of the population crosses from near-categorical use of one to near-categorical use of the other variant. A first striking finding when analysing the simulation results is that changes are rare: of the $120,960$ simulation runs using the momentum bias, only $18,040$ (around 15%) ever exhibit a directed transition, while the majority of runs simply converge on categorical use of the majority variant. This result is in line with the observation that the actuation of language change is *sporadic*: even when a novel variant is known to the entire population, this alone is not likely to lead to a community-wide language change.

[Figure 4 about here.]

When we investigate the transitions across the different parameter settings, we find that the bias strength b carves the space into two regions with distinct dynamics: while simulation runs with $b \geq 1$ exhibit directed transitions at comparable time scales, the neutral evolution condition with $b = 0$ as well as the weak momentum bias setting at $b = .5$ yield both fewer and temporally less consistent transitions, as shown in Fig. 4. The difference between those two regimes is exacerbated as population sizes become larger, making transitions in the neutral evolution conditions even rarer and slower.

Beyond this qualitative difference in successful transitions, our earlier prediction regarding the general directedness of trajectories in the neutral evolution condition are also borne out by the simulations: of all simulation runs where the incoming variant ever reaches the half-way mark (average 50% usage across the population), only 55% of trajectories in conditions with $b \leq .5$ actually result in the diffusion of the incoming variant, while the other half of the trajectories revert back to the established variant,

[5] More than 99% of simulation runs had terminated before this time limit was reached.

representing interrupted changes. In contrast, in conditions with $b \geq 1$, 97% of the trajectories that reach the half-way mark eventually lead to the population-wide adoption of the incoming variant.

In contrast to the low-bias conditions which exhibit the dynamics of neutral evolution, conditions with a sufficiently high momentum bias b will, once a change is actuated, produce reliable s-shaped transitions between the two regions of near-categorical use. The dynamics are robust under many different parameter settings which give rise to highly similar transition dynamics (see Fig. 4; the parameters' much greater influence on the likelihood of transitions occurring will be explored in a later paper). While similar transitions are also found in models driven by replicator selection, an important difference is that our model has no a priori preference for any of the variants built in. Instead of having a constant bias applied from outwith the model, the momentum term provides the opportunity for a bias to emerge dynamically and gradually from within the system, as can be seen from the temporal development of the momentum term in Figs. 5. Crucially, rather than relying on an external trigger, the s-shaped transitions are *self-actuating*: agents constantly read weak trends into the random fluctuations in their input but, across the population, these temporary individual biases will vary across the population, and more often than not cancel each other out. There is, however, always the possibility that these weak biases will overlap, causing a subset of agents to slowly shift their variant use in parallel. When this shift is detected by other agents they will themselves start to amplify it, leading to a self-reinforcing feedback loop. The directed transitions in a momentum-based model of language change are triggered *spontaneously* and, while likely, changes are not guaranteed to succeed either: even if a change is actuated, its propagation is not completely inevitable, as can be seen in interrupted changes such as the one shown in Fig. 5b. The dynamics exhibited by momentum-based selection offer an intriguing explanation of the unpredictability of the actuation of linguistic changes, by exhibiting sporadic directed changes without the need for an external bias or trigger.

[Figure 5 about here.]

The trajectories shown in Figs. 5 are exemplary of the dynamics of momentum-based selection across the full range of parameter settings we explored. Only for settings of the momentum bias b close to 0 as well as for short-term decay rates γ very close to the learning rate α do the momentum-based selection dynamics break down, and the model reverts to pure neutral evolution-like behaviour. In comparison to the prediction-driven model of Mitchener (2011), the momentum-based selection model shows that it is not necessary for learners to engage in active prediction of the population's *future* state. Rather, having a simple bias based on variant history is sufficient to drive orderly directed changes, and the transitions generated by our model appear to exhibit a more gradual uptake than the trajectories reported by Mitchener. We also find that having a bias for *regularisation* is not necessary to guarantee an orderly progression of the changes. In a population of agents who are continuously updating their usage rates, the momentum bias presented here is robust enough to drive changes to near-completion.

4 Discussion

We have shown that the momentum-based selection model fulfills two defining requirements of a model of language change: the spontaneous, sporadic actuation of changes, and their progression in the form of a directed, s-shaped curve. However, other accounts of language change which posit a selection bias in favour of the incoming variant also predict s-shaped trajectories, so how can we know which account best describes the empirical data? While the progression of every instance of language change will be influenced by several factors concurrently or at different times (see e.g. Ghanbarnejad et al., 2014; Stanford, 2014; Bickel, 2015), it is still interesting to investigate which (if any) of the mechanisms of language change discussed in the introduction can be identified as the main driving force behind language change. Here, we want to highlight some of the more subtle differences in the predictions made by different accounts of language change which would allow us to tease apart the momentum-based, language-internal and social accounts of language change based on cross-linguistic data.

4.1 The two rates of linguistic change

An interesting (and to our knowledge novel) way to evaluate competing theories of language change is to look at the predictions they make regarding the *rates* of linguistic change. It is important to note that *rate* can refer to two different things in the context of language change: one interpretation of rate is essentially the probability of a particular change occurring, such as when talking about different English past tense forms becoming regularised over time (Lieberman et al., 2007) or the rate of lexical replacement more generally (Monaghan, 2014). Rather than referring to the time frame within which a specific change takes place this is really the *likelihood of a (type of) change*, or an *actuation probability.* The other use of rate refers to the *speed* of the transition of one particular change, i.e. the time span from the introduction of a new variant to its completely replacing an older one. Under the assumption that language change follows an s-shaped pattern, this second rate of change is often taken to be the growth rate parameter of the logistic function (Pintzuk, 2003), and it is this 'rate' that is referred to by the 'Constant Rate Effect' observed in syntactic change (Kroch, 1989).

What is interesting about these two rates of change is that different accounts of language change make different predictions on whether they are correlated, i.e. whether the likelihood of a change occurring is correlated with the rate at which the change proceeds once it has been actuated. Assuming that the same pressures that lead to the introduction of more functional or 'adaptive' variants are also responsible for their preferred selection, language-internal accounts would predict that changes which occur more often cross-linguistically should also be selected for more strongly in individual languages. This would translate into faster changes so that, controlling for other factors such as frequency, the two rates of change should be positively correlated according to language-internal accounts. This differs from the prediction made by the momentum-based account: while the probability of a new variant appearing, and consequently its random actuation from the pool of variants, is dependent on linguistic factors, these factors are not what drives the selection of the variant. Assuming that individuals apply similar momentum biases to all linguistic variables, a momentum-based account would predict the speed of individual transitions to be uncorrelated with the changes' actuation probability.

The situation with social accounts is trickier: the fact that many different social factors have been posited to influence the selection of a variant, both positively and negatively, makes it difficult to derive a general prediction regarding the speed of individual changes. What determines the probability of actuation is an equally open question: it has been proposed that the actuation of changes might be driven by the need to create distinct social identities within a community (Labov, 2002; Matthews et al., 2012; Roberts, 2013), meaning that we should not expect actuation probabilities to be constant cross-linguistically. While it is difficult to pin down the exact predictions made by social accounts of language change, the language-internal and momentum-based accounts can be tested by investigating the correlation between the two rates of change that are attested cross-linguistically.

4.2 Momentum-sensitivity in the individual

While our model successfully reproduces the macro-level s-shaped curves that are characteristic of linguistic change, this raises the question of whether it makes valid assumptions about individuals' micro-level behaviour (Mesoudi and Lycett, 2009). Firstly, it is clear that both linguistic knowledge and performance are embedded in diachrony – language users are sensitive to changes in the frequencies of variants (Jaeger and Snider, 2013) and well aware of diachronic connotations (Labov, 2001; Walker and Hay, 2011; Tagliamonte, 2012), both types of information that could drive momentum-based selection. In the general cultural evolution literature it is well-established that frequency-dependent biases are a natural strategy for social learning tasks, since frequency can be an indicator of the *social value* of a variant (Boyd and Richerson, 1985). Similarly, changes in frequency can be a good indicator of the *future* social value of a cultural variant (Gureckis and Goldstone, 2009), and laboratory experiments on cultural evolution in humans have provided empirical evidence of the self-perpetuating nature of trends, where people will amplify trends even against their own personal preferences (Salganik and Watts, 2008; Willer et al., 2009). Even though this suggests that individuals would have an incentive to use metalinguistic information about the history of linguistic variants, evidence regarding the extent of people's explicit or implicit knowledge about ongoing changes is mostly qualitative and anecdotal (see e.g. Trudgill (1972); Labov (2001); Guy (2003); Tagliamonte (2012)). While variationist linguists customarily uncover patterns in the age distribution of linguistic variation based on collected data, it remains to be tested quantitatively how well (and by what mechanisms) individual speakers are capable of detecting such patterns *in the wild*.

5 Conclusion

To conclude, in this paper we investigated a new mechanism for the selection of cultural traits and studied its evolutionary dynamics, with a particular focus on the domain of linguistic change. Our analysis shows that the momentum-based selection model – where individuals are biased towards variants which have recently seen an increase in their frequency of use – fulfills two characteristic requirements of a model of language change: the spontaneous, sporadic actuation of changes, and their progression in the form of an s-shaped curve. We highlighted a number of open empirical questions related to both population-level patterns as well as the under-studied capacity of individuals to detect ongoing changes which need to be tackled in order to allow us to distinguish different accounts of language change.

References

Acerbi, A., Ghirlanda, S., and Enquist, M. (2012). The logic of fashion cycles. *PloS one*, 7(3):e32541.

Altmann, G., von Buttlar, H., Rott, W., and Strauß, U. (1983). A law of change in language. In Brainerd, B., editor, *Historical linguistics*, pages 104–115. Studienverlag Dr. N. Brockmeyer, Bochum.

Bailey, C.-J. N. (1973). *Variation and linguistic theory.* Center for Applied Linguistics, Arlington, VA.

Baxter, G. J., Blythe, R. A., Croft, W., and McKane, A. J. (2006). Utterance selection model of language change. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, 73(4 Pt 2).

Baxter, G. J., Blythe, R. A., Croft, W., and McKane, A. J. (2009). Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, 21(02):257.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., and Schoenemann, T. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, 59(December):1–26.

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

Bentley, R. A., Lipo, C. P., Herzog, H. A., and Hahn, M. W. (2007). Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior*, 28(3):151–158.

Berger, J. and Le Mens, G. (2009). How adoption speed affects the abandonment of cultural tastes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(20):8146–8150.

Bickel, B. (2015). Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In Heine, B. and Narrog, H., editors, *The*

Oxford handbook of linguistic analysis, pages 901–923. Oxford University Press, Oxford.

Blythe, R. A. (2012). Neutral evolution: a null model for language dynamics. *Advances in Complex Systems*, 15(03n04):1150015.

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

Bowie, D. and Yaeger-Dror, M. (2013). Phonological Change in Real Time. In Salmons, J. and Honeybone, P., editors, *The Oxford handbook of historical phonology*, chapter 34. Oxford University Press, Oxford.

Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.

Bybee, J. (2007). *Frequency of Use and the Organization of Language*. Oxford University Press.

Croft, W. (2000). *Explaining language change: an evolutionary approach*. Pearson Education Limited, Harlow.

Croft, W. (2006). The relevance of an evolutionary model to historical linguistics. In Thomsen, O. N., editor, *Different models of linguistic change*, Current Issues in Linguistic Theory, pages 91–132. John Benjamins Publishing Company.

de Saussure, F. (1959). *Course in general linguistics*. The Philosophical Library, Inc., New York.

Denison, D. (2003). Log(ist)ic and simplistic S-curves. In Hickey, R., editor, *Motives for language change*, pages 54–70. Cambridge University Press, Cambridge.

Ellegard, A. (1953). *The auxiliary do: the establishment and regulation of its use in English*. Gothenburg studies in English. Almqvist & Wiksell, Stockholm.

Evans, N. and Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *The Behavioral and brain sciences*, 32(5):429–48; discussion 448–494.

Foulkes, P. and Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34(4):409–438.

Foulkes, P. and Vihman, M. (2013). First language acquisition and phonological change. In Honeybone, P. and Salmons, J., editors, *The Oxford handbook of historical phonology*, chapter 18. Oxford University Press, Oxford.

Ghanbarnejad, F., Gerlach, M., Miotto, J. M., and Altmann, E. G. (2014). Extracting information from S-curves of language change. *Journal of the Royal Society, Interface*, 11(101):20141044.

Greenberg, J. H. (1959). Language and evolution. In Meggers, J., editor, *Evolution and anthropology: a centennial appraisal*. Washington DC.

Griffiths, T. L. and Kalish, M. L. (2007). Language Evolution by Iterated Learning with Bayesian Agents. *Cognitive Science*, 31(3):441–480.

Gureckis, T. M. and Goldstone, R. L. (2009). How You Named Your Child: Understanding the Relationship Between Individual Decision Making and Collective Outcomes. *Topics in Cognitive Science*, 1(4):651–674.

Guy, G. R. (2003). Variationist Approaches to Phonological Change. In Joseph, B. D. and Janda, R. D., editors, *The Handbook of Historical Linguistics*, chapter 8, pages 369–400. Blackwell Publishing Ltd, Oxford, UK.

Hockett, C. F. (1958). *A course in modern linguistics*. The Macmillan Company, New York.

Hockett, C. F. (1965). Sound change. *Language*, 41(2):185–204.

Jaeger, T. F. and Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127(1):57–83.

Jaeger, T. F. and Tily, H. (2010). On language utility: processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3):323–335.

Jespersen, O. (1922). *Language: Its Nature, Development and Origin*. George Allen & Unwin Ltd., London.

Kerswill, P. (1996). Children, adolescents, and language change. *Language Variation and Change*, 8(2):177–202.

Kiparsky, P. (1968). Linguistic universals and linguistic change. In Bach, E. and Harms, R. T., editors, *Universals in linguistic theory*, pages 170–202. Holt, Rinehart, and Winston, New York.

Kirby, S. (1999). *Function, Selection, and Innateness.* Oxford University Press, Oxford.

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

Kroch, A. S. (1994). Morphosyntactic variation. In Beals, K., editor, *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory.*

Kroeber, A. L. (1919). On the Principle of Order in Civilization as Exemplified by Changes of Fashion. *American Anthropologist*, 21(3):235–263.

Labov, W. (1994). *Principles of linguistic change. Internal factors*, volume 1 of *Language in Society.* Blackwell, Oxford.

Labov, W. (2001). *Principles of linguistic change. Social factors*, volume 2. Blackwell Publishers Inc., Malden, Massachusetts.

Labov, W. (2002). Driving forces in linguistic change. In *2002 International Conference on Korean Linguistics*, Seoul. Seoul National University.

Lass, R. (1980). *On explaining language change.* Cambridge University Press, Cambridge.

Lewis, H. M. and Laland, K. N. (2012). Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599):2171–2180.

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

Lightfoot, D. (1979). *Principles of diachronic syntax.* Cambridge University Press, Cambridge.

Lightfoot, D. (1991). *How to set parameters: Arguments from language change.* MIT Press, Cambridge, Massachusetts.

Matthews, C., Roberts, G., and Caldwell, C. A. (2012). Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. *Evolution and Human Behavior*, 33(6):759–770.

McMahon, A. M. S. (1994). *Understanding language change.* Cambridge University Press, Cambridge.

Meillet, A. (1921). *Linguistique historique et linguistique generale.* La societie linguistique de Paris, Paris.

Mesoudi, A. (2011). *Cultural Evolution.* University of Chicago Press, Chicago.

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

Mitchener, W. G. (2011). A mathematical model of prediction-driven instability: how social structure can drive language change. *Journal of Logic, Language and Information*, 20:385–396.

Monaghan, P. (2014). Age of acquisition predicts rate of lexical evolution. *Cognition*, 133(3):530–534.

Nardy, A., Chevrot, J.-P., and Barbu, S. (2013). The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics*, 51(2):255–284.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In Breivik, L. E. and Jahr, E. H., editors, *Language change: contributions to the study of its causes*, pages 173–198. Mouton de Gruyter, Berlin.

Ohala, J. J. (1993). The phonetics of sound change. In Jones, C., editor, *Historical linguistics: problems and perspectives*, pages 237–278. Longman, London.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In Gussenhoven, C. and Warner, N., editors, *Laboratory Phonology VII*, volume 7 of *Phonology & Phonetics*, pages 101–140. Mouton de Gruyter, Berlin.

Pintzuk, S. (2003). Variationist approaches to syntactic change. In Joseph, B. D. and Janda, R. D., editors, *The Handbook of Historical Linguistics*, chapter 15, pages 509–528. Blackwell Publishing Ltd, Oxford.

Postal, P. M. (1968). *Aspects of phonological theory.* Harper & Row, New York.

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

Roberts, G. (2013). Perspectives on Language as a Source of Social Markers. *Language and Linguistics Compass*, 7(12):619–632.

Salganik, M. J. and Watts, D. J. (2008). Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market. *Social Psychology Quarterly*, 71(4):338–355.

Sankoff, D. (1988). Sociolinguistics and syntactic variation. In Newmeyer, F. J., editor, *Linguistics: The Cambridge Survey*, chapter 8, pages 140–161. Cambridge University Press, Cambridge.

Sankoff, G. and Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, 83(3):560–588.

Stanford, J. N. (2014). Language acquisition and language change. In Bowern, C. and Evans, B., editors, *The Routledge Handbook of Historical Linguistics*, chapter 21, pages 466–483. Routledge, London.

Steels, L. (2000). Language as a Complex Adaptive System. In Schoenauer, M., Deb, K., Rudolph, G., Yao, X., Lutton, E., Merelo, J. J., and Schwefel, H.-P., editors, *Parallel Problem Solving from Nature PPSN VI*, pages 17–26. Springer.

Stevens, M. and Harrington, J. (2013). The individual and the actuation of sound change. *Loquens*, 1(1):e003.

Sturtevant, E. H. (1947). *An introduction to linguistic science.* Yale University Press, New Haven.

Swarup, S. and McCarthy, C. (2012). Representational Momentum May Explain Aspects of Vowel Shifts. In *Artificial Life 13*, pages 267–274. MIT Press.

Tagliamonte, S. A. (2012). *Variationist sociolinguistics. Change, observation, interpretation.* Language in Society. Wiley-Blackwell.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, 1(02):179–195.

Trudgill, P. (2004). *New-dialect formation: the inevitability of Colonial Englishes*. Edinburgh University Press, Edinburgh.

Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society*, 37(02):241–254.

Vennemann, T. (1983). Causality in language change: theories of linguistic preferences as a basis for linguistic explanations. *Folia Linguistica Historica*, VI(1):5–26.

Walker, A. and Hay, J. B. (2011). Congruence between word age and voice age facilitates lexical access. *Laboratory Phonology*, 2(1):219–237.

Wang, W. S.-Y. (1969). Competing Changes as a Cause of Residue. *Language*, 45(1):9–25.

Wedel, A., Kaplan, A., and Jackson, S. (2013). High functional load inhibits phonological contrast loss: a corpus study. *Cognition*, 128(2):179–186.

Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, 23(3):247–274.

Weinreich, U., Labov, W., and Herzog, M. (1968). Empirical foundations for a theory of language change. In Lehmann, W. P. and Malkiel, Y., editors, *Directions for Historical Linguistics: A Symposium*, pages 95–188. University of Texas Press.

Willer, R., Kuwabara, K., and Macy, M. W. (2009). The False Enforcement of Unpopular Norms. *American Journal of Sociology*, 115(2):451–490.

Winter-Froemel, E. (2008). Towards a comprehensive view of language change: Three recent evolutionary approaches. In Detges, U. and Waltereit, R., editors, *The Paradox of Grammatical Change: Perspectives from Romance*, volume 293 of *Current Issues in Linguistic Theory*, pages 215–250. John Benjamins Publishing Company, Amsterdam.

Figures

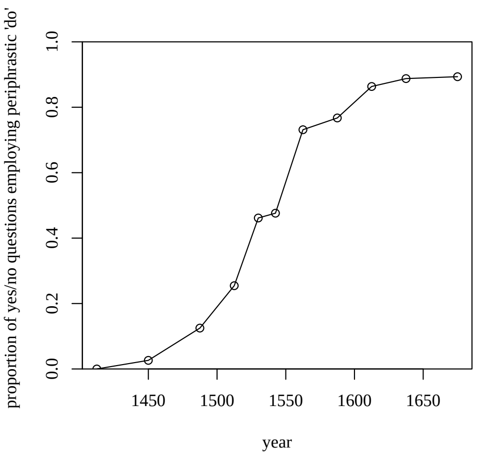

Figure 1: Competition between two syntactic patterns of *yes/no questions*, as observed in a corpus of Middle English writing (Ellegard, 1953). The established question syntax (e.g. "Went he?") was gradually replaced by its modern variant (e.g. "Did he go?") along an s-shaped trajectory.

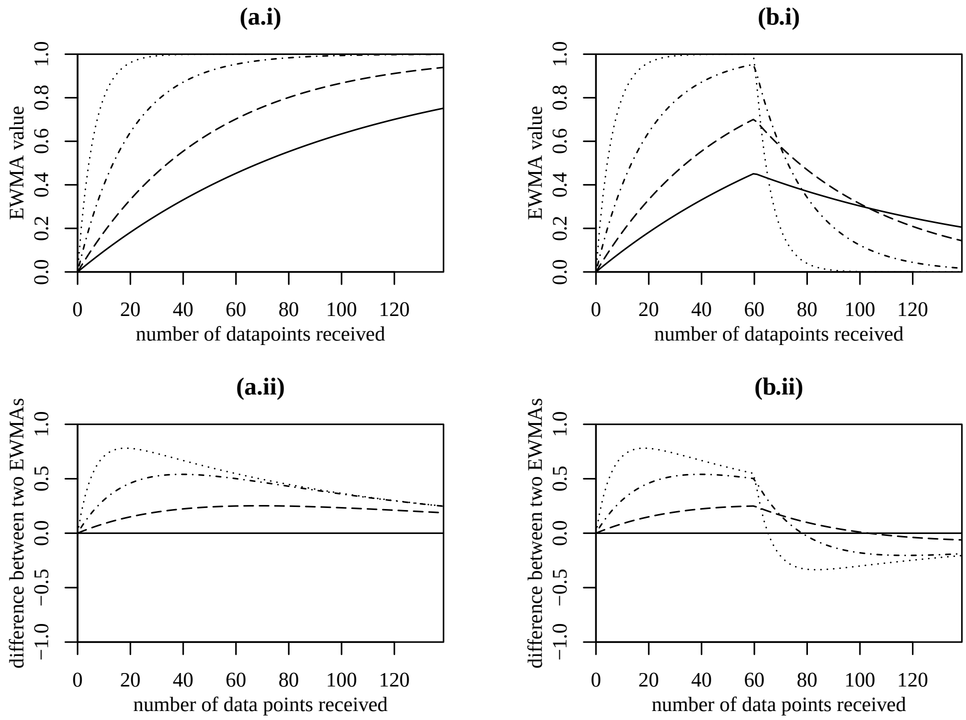

Figure 2: Illustration of how the interaction between exponentially weighted moving averages (EWMAs) of the same input data but with different decay parameters (upper graphs) affects the temporal development of the corresponding momentum terms (lower graphs). **(a)** Four EWMAs with decay rates $\gamma = .01, .02, .05, .15$ are initialised at $\hat{n}_\gamma = 0$ and repeatedly updated using the same constant input data series $\vec{n} = \langle 1, 1, 1 \ldots \rangle$. *(i)* The higher the decay parameter, the faster the EWMAs approach the input values; the slowest (solid) line shows the development of the EWMA with $\gamma = .01$, the fastest (dotted) line $\gamma = .15$. *(ii)* Corresponding momentum terms $m(t) = \hat{n}_\gamma(t) - \hat{n}_\alpha(t)$ derived from the EWMAs above, by taking each of the EWMAs and subtracting the value of the EWMA with the slowest decay rate $\alpha = .01$ (line styles correspond to those in (i)). A value of γ further away from α decreases the time t_{mmax} until the maximally possible momentum is reached while making the overall time-course of momentum more peaky, with a higher maximum value $m_{\max}$ and quicker decay back towards 0 following the peak. **(b)** Same as (a), only that the EWMAs' input data series $\vec{n}$ switches from all 1s to all 0s after 60 data points. *(i)* The EWMAs with the highest decay parameter quickly converge back towards the new input target 0. *(ii)* Corresponding momentum terms derived from the EWMAs above, again subtracting the value of the EWMA with the slowest decay rate $\alpha = .01$ (line styles correspond to those in (i)). The sudden change in trend after 60 data points illustrates how the two parameters α, γ control the time depth at which the momentum term is most sensitive to underlying trends in the data: momentum terms based on high γ (e.g. $\gamma = .15$, dotted line), while very quick to reflect sudden changes in the input, are very unstable. After five data points indicating a new downward trend back towards 0, the previous sustained upward trend is forgotten, with the momentum term quickly returning to 0, then going negative to reflect the new downward trend. Momentum terms based on settings of γ closer to α (e.g. $\gamma = .02$, dashed line) are more conservative, requiring sustained evidence of a trend over time to reach a high value.

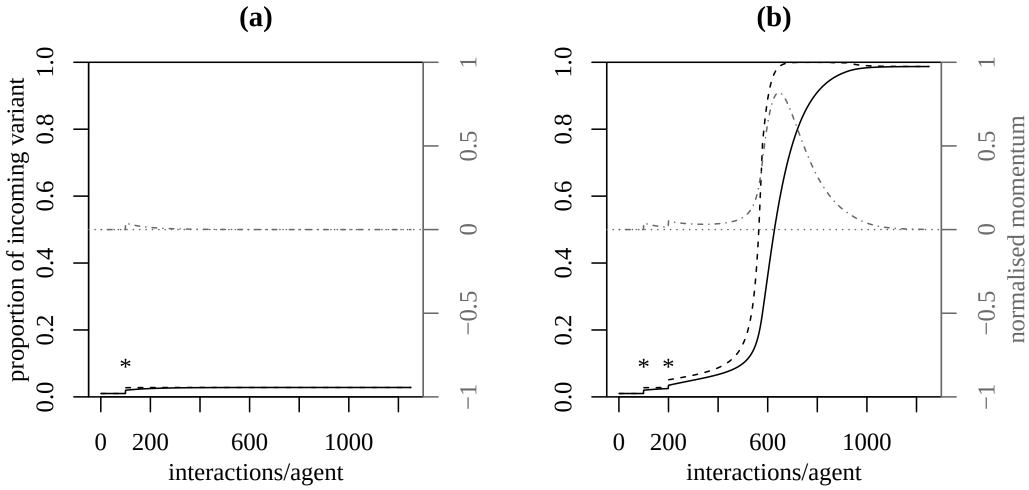

Figure 3: Momentum-based selection dynamics of a single agent's variable usage rate in a production-perception loop, with learning rates $\alpha = 0.01, \gamma = 0.02$ and momentum bias $b = 2$. At every time step the agent updates their own usage rate (solid black line) by aligning to their own average momentum-biased production with a sample resolution of $T = 5$ (indicated by the dashed black line). This stable loop is perturbed by administering fabricated input data suggesting 100% usage of the incoming variant at the time points marked by asterisks, demonstrating the two regimes of momentum-based selection:
(a) stability: a single fabricated data point after 100 interactions causes a sudden increase in the agent's usage rate (solid black line) as well as the momentum term (dot-dashed grey line, right axis). The positive momentum term causes the agent's own perceived usage level to be higher than it actually is (dashed black line), which leads to some further increase in the usage rate before the momentum bias tapers off towards 0 (the feedback loop stabilises again after around 500 interactions). **(b)** directed transitions: adding another fabricated data point after 200 interactions raises the momentum term high enough to trigger self-reinforcing runaway change, giving rise to an s-shaped transition.

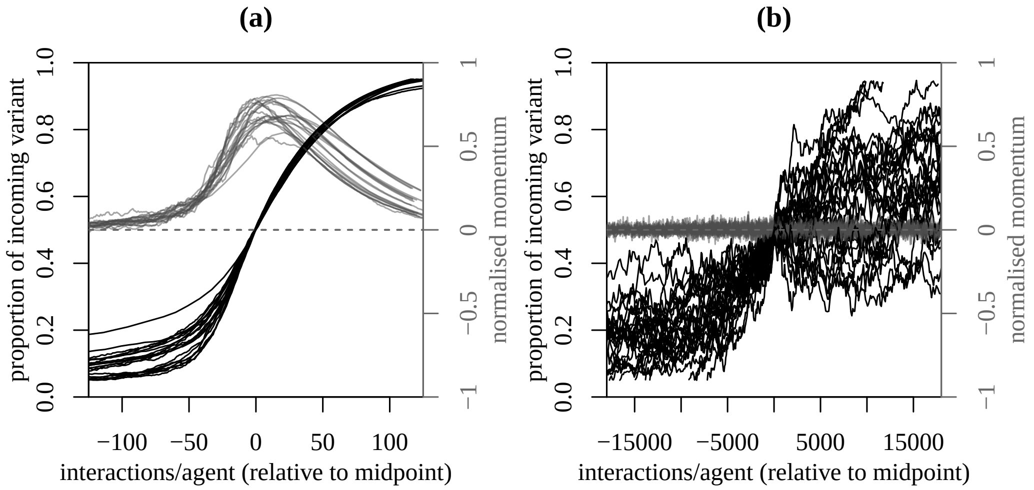

Figure 4: Successful transitions generated by simulation runs in conditions with and without the momentum-based selection bias. The graphs show the development of the average proportion of use of the incoming variant across the population (black line, left axis) from the point where it crosses the 5% mark until it reaches 95%, alongside the average momentum term during that period (grey line, right axis). Transitions are aligned at the point where the trajectory first crosses the 50% mark of incoming variant usage. **(a)** 20 trajectories randomly drawn from the $21,909$ successful transitions generated by momentum-based selection with momentum bias $b \geq 1$, population sizes $N \geq 5$ and various settings of γ, T, x_0. The momentum term influences the agents' perception of the usage levels around them which, once triggered, leads to a self-reinforcing feedback loop. **(b)** all 28 transitions generated in $17,280$ simulation runs with $b = 0$, equivalent to neutral evolution, with various settings of γ, T, x_0 and population sizes $N \geq 5$. Without the influence of the momentum bias, transitions become both much rarer and slower as population size increases (note the different time scales). The momentum term, ineffective in this model, is shown for reference.

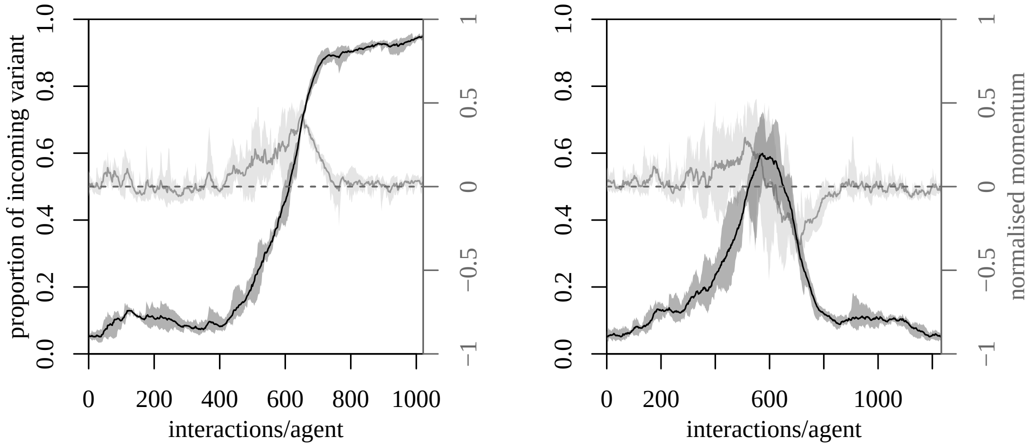

Figure 5: Transitions generated by two simulation runs using identical parameter settings ($N = 5, b = 2.0, T = 2, \alpha = .01, \gamma = .04$). The graphs show the development of the average proportion of use of the incoming variant across the population (black line, left axis) as well as the average momentum term influencing the agents' perception (grey line, right axis). Shaded intervals indicate the range (minimum and maximum values) attested in the population. **(a)** A successful, s-shaped transition typical of momentum-based selection: an initially noisy momentum value rises high enough to trigger self-reinforcement of the momentum bias (at around 450 interactions) until it saturates and tails off again **(b)** Example of a rare, interrupted transition: despite the onset of a directed shift, the wide range of momentum biases across the population destabilises the feedback loop, causing the average momentum to break down and invert, returning the usage frequency of the incoming variant back towards its initial low level.