Original Articles

Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model

Simon Todd[a,*], Janet B. Pierrehumbert[b,c], Jennifer Hay[c,d]

[a] Department of Linguistics, Stanford University, Margaret Jacks Hall, Building 460, Stanford, CA 94305-2150, United States
[b] Oxford e-Research Centre, University of Oxford, 7 Keble Road, Oxford OX1 3QG, United Kingdom
[c] New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
[d] Department of Linguistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

ABSTRACT

Empirically-observed word frequency effects in regular sound change present a puzzle: how can high-frequency words change faster than low-frequency words in some cases, slower in other cases, and at the same rate in yet other cases? We argue that this puzzle can be answered by giving substantial weight to the role of the listener. We present an exemplar-based computational model of regular sound change in which the listener plays a large role, and we demonstrate that it generates sound changes with properties and word frequency effects seen in corpora. In particular, we consider the experimentally-supported assumption that high-frequency words may be more robustly recognized than low-frequency words in the face of acoustic ambiguity. We show that this assumption allows high-frequency words to change at the same rate as low-frequency words when a phoneme category moves without encroaching on the acoustic space of another, faster than low-frequency words when it moves toward another, and slower than low-frequency words when it moves away from another. We discuss how these predicted word frequency effects apply to different types of sound changes that have been observed in the literature. Importantly, these frequency effects follow from assumptions regarding processes in perception, not production. Frequency-based asymmetries in perception predict different frequency effects for different kinds of sound change.

1. Introduction

Language, as a system of communication, relies upon both speakers and listeners. People spend more time listening than speaking (e.g. Emanuel et al., 2008, & references therein), and listeners bring an array of perceptual biases to the understanding of speech (e.g. Connine, Titone, & Wang, 1993; Ganong, 1980; Hay, Warren, & Drager, 2006; Hay & Drager, 2010; Kleinschmidt & Jaeger, 2015; Niedzielski, 1999; Pitt & Samuel, 1993; Strand & Johnson, 1996). Moreover, speech that is listened to is stored in memory (e.g. Goldinger, 1996) and can affect future speech production and perception (see e.g. for production: Fowler, Brown, Sabadini, & Weihing, 2003; Goldinger, 1998; Nielsen, 2011; Pardo, 2006; and for perception: Bradlow & Bent, 2008; Clarke-Davidson, Luce, & Sawusch, 2008; Dahan, Drucker, & Scarborough, 2008; Kraljic & Samuel, 2006; Norris, McQueen, & Cutler, 2003). Listening is a crucial component of in-the-moment communication. It is therefore plausible that patterns of sound change may be substantially shaped by the listener. However, most theories of sound change are driven by the role of the speaker, even if they acknowledge the relevance of the listener.

Functionalist linguists have long identified the relevance of perceptual constraints to linguistic sound systems (e.g. Flemming, 2004; Liljencrants & Lindblom, 1972; Martinet, 1952). They have typically done this by taking into account the speaker's awareness of the potential difficulties they may cause for the listener; the speaker knows that using phonetically ambiguous sounds could cause communication to fail, and thus tends to avoid them (e.g. Buz, Tanenhaus, & Jaeger, 2016; Lindblom, 1990). While such listener-oriented, speaker-based approaches insightfully highlight the relevance of the listener, they may overpredict the speaker's avoidance of ambiguity (e.g. Arnold, Wasow, Asudeh, & Alrenga, 2004; Bard et al., 2000), and they have been criticized as teleological (e.g. Wedel, 2006).

An alternative approach, which we take here, is to acknowledge that listeners are also speakers, and thus that any change to cognitive representations of sound categories due to passive perceptual processes in the listener will be reflected in the speech of that listener-turned-

---

speaker. This approach was originally presented by Ohala (1981) with a model of categorical change in particular words via misperception of sounds in their phonological environments. It can be extended to model gradient change in all words via biases involved in the correct perception of spoken words, as illustrated by Harrington, Kleber, Reubold, Schiel, and Stevens (2018). We implement such an extension, drawing on gradient perceptual robustness rather than categorical misperception as a driving mechanism, and placing primary focus on the lexical rather than phonological context of sounds under perceptual scrutiny (following a school of thought dating back to Courtenay, 1895).

We focus on a particular kind of change, *regular sound change*, defined as the gradual transformation of the phonetic realization of a phoneme over time (Labov, 2010). We develop a computational model that assumes that the listener plays a large role in regular sound change, drawing on experimentally-supported perceptual processes. We show that our model successfully captures empirically-attested patterns that cannot be readily explained via emphasis solely on the speaker. In particular, we focus on the rate[1] at which words of different frequencies participate in regular sound change, for which different approaches to sound change make different predictions.

According to the Neogrammarian hypothesis (see e.g. Garrett, 2015), regular sound change affects all eligible words in the same way – and thus at the same rate, regardless of frequency. This lexical independence in sound change follows from the assumption of *strict modularity*, where the representation of the phoneme is independent of its instantiation in words. Under this assumption, regular sound change involves changes to the phoneme representation rather than to words directly.

By contrast, recent usage-based approaches relax the assumption of strict modularity, contending that the instantiation of the phoneme within words is central to the way that the phoneme is represented both cognitively and theoretically (e.g. Beckner et al., 2009; Blevins & Wedel, 2009; Bybee, 2002; Harrington et al., 2018; Hay, Pierrehumbert, Walker, & LaShell, 2015; Hay & Foulkes, 2016; Johnson, 1997; Pierrehumbert, 2001, 2002; Wedel, 2006, 2012). Such approaches assume that the instances of the same phoneme in different words may have different (but related) representational bases. Consequently, while sound change is expected to affect all words containing the changing phoneme over a certain period of time, it is not assumed to affect all words at the same rate.

In particular, the Frequency Actuation Hypothesis (henceforth, FAH; Bybee, 2002; Phillips, 1984) claims that word frequency effects[2] will be different in different kinds of sound change, depending on the motivation of the change. Phillips (1984) presents a two-way distinction between physiologically motivated changes and non-physiologically motivated changes. Physiologically motivated changes result from the iteration of articulatory biases and affect the surface phonetic form of phonological segments. An example is /t/-tapping, where a word like *matter* comes to sound more like *madder*, reducing articulatory effort. In physiologically motivated sound changes, high-frequency words are predicted to change faster than low-frequency words, since they are produced and thus subjected to the articulatory bias more often. Non-physiologically motivated changes result from lexical analogy of sound

patterns to new environments and yield new constraints on underlying phonological or phonotactic structures. An example is the deletion of glides after coronal stops /t d n/, where a word like *tune* comes to sound more like *toon*, generalizing the constraint banning glides after other coronal consonants (Phillips, 1981). In non-physiologically motivated sound changes, high-frequency words are predicted to change more slowly than low-frequency words, since their frequent use allows them to persist as exceptions in the phonological grammar.

The most intuitive application of the FAH to regular sound change makes the assumption that gradient phonetic change results primarily from iterated biases in the speaker's phonetic implementation, and thus predicts that high frequency words should always change faster than low-frequency words. This prediction found support in early usage-based modeling work by Pierrehumbert (2001), in which the speaker was central. However, it does not hold uniformly in empirical data. We are aware of three corpus studies of word frequency effects on rates of sound change across the lexicon.[3] One of these studies found a result that is inconsistent with predictions of the FAH while not directly opposing them: Bermúdez-Otero, Baranowski, Bailey, and Turton (2015) found that /t/-glottaling in Manchester English is affecting words of all frequencies *at the same rate*. Another found a result that is fully consistent with predictions of the FAH: Hay and Foulkes (2016) found that /t/-tapping in New Zealand English is affecting high-frequency words faster than low-frequency words. The final study, however, found a result that opposes predictions of the FAH: Hay et al. (2015) found that /ɛ/-raising (and other processes in the same vowel shift) in New Zealand English affected high-frequency words *slower* than low-frequency words. We discuss these three changes in more detail in Section 2.1; for now, we simply note that the FAH does not give considerable reason to treat them differently. The existence of different effects of word frequency on rate of change in different kinds of change remains an unsolved puzzle in studies of regular sound change.

We propose that these differences can be understood by making the listener central to regular sound change. We hypothesize that asymmetries in the rates with which regular sound change affects different words follow from experimentally-supported asymmetries in the robustness with which those words can be recognized: high-frequency words can be recognized more robustly than low-frequency words in the face of acoustic ambiguity. We further hypothesize that different asymmetries are observed in different kinds of regular sound change because of the different implications they have for the acoustic ambiguity of the involved phonemes. High-frequency words change at the same rate as low-frequency words when a phoneme moves without encroaching on the acoustic space of another, with no bearing on acoustic ambiguity; faster than low-frequency words when a phoneme moves toward another, potentially increasing acoustic ambiguity; and slower than low-frequency words when a phoneme moves away from another, potentially decreasing acoustic ambiguity. This paper develops a computational model that allows us to test these hypotheses and to derive general predictions for the effects of word frequency in other cases of regular sound change.

The remainder of this paper is structured as follows. In Section 2, we give an overview of the three existing corpus-based investigations of word frequency effects on rate of change, the general considerations we take away for modeling such changes, and the framework in which our model is based. In Section 3, we lay out the basic model and discuss how it generates sound change. In Section 4, we show how the model generates sound changes with implications for just a single phoneme,

---

[1] Traditionally, linguists have considered the effects of word frequency on the *actuation* of sound change – i.e. whether high- or low-frequency words change first. We focus instead on *rates* of change – i.e. whether high- or low-frequency words change fastest – because they can be identified more easily and robustly in corpora without extensive time depth (as a statistical interaction between word frequency and time), and because they allow for easier disentanglement of change from natural phonetic variation in a continuous acoustic system.

[2] The FAH was originally posited with reference to the question of whether high- or low-frequency words change first. For the reasons previously outlined, we reinterpret it to generate predictions about whether high- or low-frequency words change fastest, under the assumption that change begins from neutral initial conditions.

---

[3] We include only studies that aggregate across the lexicon because the assumption that speech and/or sound change contain stochastic elements implies that sampling a few words is not sufficient to reflect upon the existence of statistical tendencies tied to word frequency. For this reason, we exclude from consideration a study by Tamminga (2014), which explores /ai/-raising in Philadelphia English for various senses of *like*.

and we discuss the model's predictions of word frequency effects in this case. In Section 5, we show how the model generates sound changes with implications for two phonemes, and we show that enhancing the model with a bias for recognizing high-frequency words more robustly than low-frequency words generates word frequency effects on rate of change that are consistent with the empirical evidence. In Section 6, we discuss the implications of this result and the predictions it makes for other kinds of sound change. The paper concludes in Section 7.

2. Overview

We develop our model to show how an emphasis on the listener can predict different effects of word frequency on rate of change in different kinds of regular sound change. For concrete empirical grounding, we focus on the three existing studies of different sound changes that consider such effects, and we aim to show that experimentally-supported perceptual processes can explain the differences between them. For simplicity, we aim not to model the full details of each change, but rather to capture certain key properties of each change, both frequency-independent and frequency-based. In this section, we describe the relevant data from each change, how we choose to model the changes, the desiderata that the data yield for our model, and the general approach that must be taken by any model aiming to meet these desiderata. Finally, we introduce the specific modeling framework that we use to implement this approach, Exemplar Theory.

2.1. Empirical sound change data

Our model aims to capture key results from the three existing studies of regular sound change that consider word frequency effects on rate of change, across the lexicon: /t/-glottaling in Manchester English (Bermúdez-Otero et al., 2015), /t/-tapping in New Zealand English (Hay & Foulkes, 2016), and /ɛ/-raising in the short front vowel shift of New Zealand English (Hay et al., 2015).

/t/-glottaling refers to a sound change whereby /t/ between vowels becomes increasingly likely to be realized as a glottal stop /ʔ/[4]; an example is *mitten* coming to be pronounced as "mi'en". Bermúdez-Otero et al. (2015) present a study of /t/-glottaling in Manchester English, which finds that the use of /t/-glottaling has increased over time. Crucially, while high-frequency words exhibit more /t/-glottaling at every point of time, words of all frequencies have changed at the same *rate*.

/t/-tapping is also a sound change that affects the realization of /t/ between vowels. Under tapping, intervocalic /t/ is weakened to a shorted voiced sound that may be notated [ɾ] or [d]; an example is *matter* coming to be pronounced more like "madder". Hay and Foulkes (2016) present a study of /t/-tapping in New Zealand English, which finds that the use of /t/-tapping has increased over time. Crucially, unlike for /t/-glottaling in Manchester English, high-frequency words have exhibited this change at a faster rate than low-frequency words.

Unlike the other sound changes, /ɛ/-raising affects vowels, not consonants. It causes the vowel /ɛ/ to be pronounced with the tongue higher in the mouth, making New Zealand *bet* sound to non-New Zealanders like "bit". This process is part[5] of a *push chain*, where one phoneme (in this case, /æ/) moves toward another (/ɛ/) in acoustic

space and pushes it along a related trajectory of change. Hay et al. (2015) present a study of the New Zealand short front vowel shift, the larger push chain of which /ɛ/-raising is a component. Crucially, they find an effect of word frequency on rate of change which is different to that seen in the previous two studies: high-frequency words have changed *slower* than low-frequency words.

2.2. Modeling the sound changes

The sound changes outlined in Section 2.1 are complex, and modeling them in full detail far exceeds the scope of this paper. Instead, we aim to construct a model whose output resembles them at a high level. Here, we describe the way that we choose to model the three sound changes.

We choose to model /t/-glottaling as directed phonetic drift in the realization of a single, isolated phoneme. We make the decision to treat /t/-glottaling as isolated because it only affects /t/ and because it produces realizations that remain unlike any other phoneme of English (as there is no phoneme /ʔ/). Thus, we identify /t/-glottaling with a system containing a single phoneme category subject to a consistent production bias. We ultimately expect our model to show movement of high- and low-frequency words in this phoneme category *at the same rate*.

Unlike /t/-glottaling, /t/-tapping produces realizations that are more like those of an existing phoneme of English, /d/. Consequently, it is not sufficient to think of /t/-tapping as isolated; rather, the presence of the competing phoneme /d/ must also be acknowledged. However, since the empirical data of interest concern /t/ and not /d/, we are only fundamentally concerned with modeling the implications of the *existence* of /d/ for the changes in /t/, and not with modeling the behavior of /d/ itself. Thus, we identify /t/-tapping with a system containing two phoneme categories where one is biased toward the other, and we focus on the phoneme category subject to the bias. We ultimately expect our model to show *faster* movement of high-frequency words in this phoneme category than of low-frequency words.

The presence of competing phonemes is also relevant to /ɛ/-raising. For simplicity, we abstract from the full push chain to an interaction between just two phonemes, /æ/ and /ɛ/, along a single dimension (height). It is a consequence of this simplification that we cannot model the behavior of /æ/, only of /ɛ/ (see Appendix A.1 for discussion). Thus, we identify /ɛ/-raising with a system containing two phoneme categories where one is biased toward the other, and we focus on the phoneme category which is not subject to the bias. We ultimately expect our model to show *slower* movement of high-frequency words in this phoneme category than of low-frequency words.

For convenience, we choose to model the elements of interest from /t/-tapping together with those from /ɛ/-raising. That is, we create a single system of exemplars belonging to two interacting phonemes, and we base our expectations for the behaviors of the two different phonemes on the two different kinds of change. For ease of reference, we designate the biased phoneme category the Pusher and the other phoneme category the Pushee. A model of word frequency effects in /t/-tapping needs to focus on the Pusher, while a model of the observed word frequency effects in /ɛ/-raising needs to focus on the Pushee. Viewing the encroachment of the Pusher on the Pushee as analogous to the encroachment of /t/ on /d/ in New Zealand English /t/-tapping, we expect the model to show faster movement of high-frequency words than of low-frequency words in the Pusher. Similarly, viewing the retreat of the Pushee from the Pusher as analogous to the retreat of /ɛ/ from /æ/ in the New Zealand short front vowel shift, we expect the model to show slower movement of high-frequency words than of low-frequency words in the Pushee.

We also make the simplifying assumption of ignoring minimal pairs, i.e. words such as *bat* and *bet* which are distinguished solely by the phonemic distinction at the heart of the sound change in question. We make this choice not because we think that minimal pairs are irrelevant

---

[4] Realization of /t/ as /ʔ/ entails the removal of oral gestures. This phenomenon, also known as *glottal replacement*, is common across British varieties of English and almost exclusively affects /t/ (see e.g. Milroy, Milroy, Hartley, & Walshaw, 1994, for evidence from Tyneside English). It is articulatorily and acoustically distinct from the application of glottal constriction to oral gestures (*glottal reinforcement*), which more commonly affects /p/ and /k/.

[5] While the larger push chain has implications for multiple phoneme categories, we focus here on establishing a thorough understanding of the /ɛ/-raising component because it is least affected by model simplifications (see Appendix A.1 for discussion).

to regular sound change, but rather because we believe that they *alone* are unlikely to explain the properties that hold across all words in a phoneme category, such as frequency effects. This belief is supported by both empirical evidence and simulations, which we discuss in Appendix A.2.

Because our model is based on specific cases of sound change only at a high level, it extends beyond these cases to schematize word frequency effects in single-category movements and two-category interactions in general. Of course, some instances of regular sound change cannot be treated as single-category movement or two-category interaction, or have other properties that do not meet with our simplifying assumptions (see Appendix A.1). Nevertheless, the broad insights of our model are widely applicable and constitute a necessary foundation for understanding the dynamics of highly complex instances of regular sound change.

2.3. Model desiderata

It is clear that capturing word frequency effects on rates of change will require a phoneme to have different, but related, representational bases for different words. What other requirements do the empirical data present for our model?

Before investigating word frequency effects, we require our model to generate single-category movements and two-category interactions that resemble the empirical sound changes reviewed in Section 2.1 at a high level. Thus, we require the model to produce certain key frequency-independent properties, which we infer on the basis of the data from the New Zealand short front vowel shift.[6]

These key properties relate to the maintenance of structure over the course of change. As vowel distributions moved in the New Zealand short front vowel shift, they maintained their distance from one another, their shapes (width and skewness), and their degree of overlap with one another. At all times, they exhibited little skewness and substantial overlap relative to their widths; such properties are also seen in vowels in American English (Hillenbrand, Getty, Clark, & Wheeler, 1995). We illustrate these key properties for /æ/ and /ɛ/ over a 60-year period of the data in Fig. 1.

The basic desiderata for our model are therefore that it: (i) generates movement of each category; (ii) maintains the shape (width and skewness) of each category; and, in two-category interactions, maintains the (iii) distance between and (iv) overlap of the categories. To our knowledge, no other exemplar-based model has met all these desiderata, so this is an important way in which our model makes a basic contribution, independent of word frequency effects.

Though a computational model is the only way in which to test causal hypotheses in the study of sound change, such tests can only bear on real sound change to the extent that the model captures empirically-observable key properties of sound systems and sound change. Thus, meeting the model desiderata is a necessary precursor to our investigation of word frequency effects on rates of change.

2.4. Meeting desiderata: general approach

What must a model do to meet the desiderata from Section 2.3? A high-level demonstration can be given by considering forces that act upon the phoneme category distributions (Fig. 2).

A model which meets our desiderata must instantiate and balance a number of forces. A model of a single-category system must initiate and
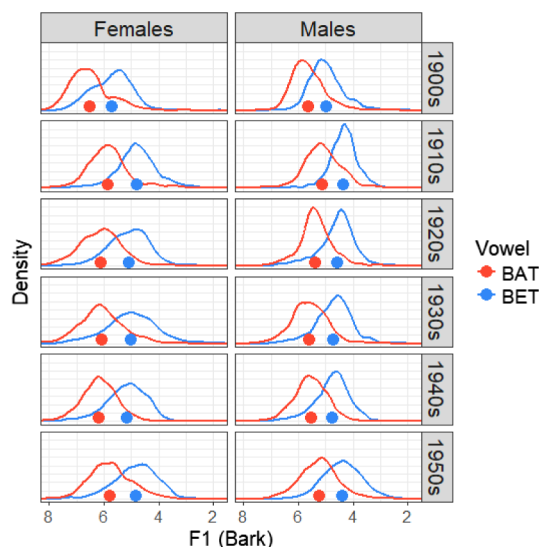
Fig. 1. Centroids (points) and distributions (densities) for the F1 values of the BAT (/æ/; red) and BET (/ɛ/; blue) vowel categories for speakers of New Zealand English born each decade from 1900 to 1959 (top to bottom), based on the raw data from Hay et al. (2015). While the category centroids move over time, their distance from one another stays approximately constant. The shapes (width and skewness) of the category distributions also stay approximately constant over time, as does the substantial degree of overlap between the two distributions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
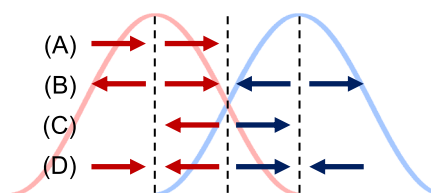
Fig. 2. Forces that minimally must act on phoneme category distributions in order to meet model desiderata. (A) An intrusive force pushes the Pusher toward the Pushee. (B) A spreading force allows the far side of the Pushee (blue) to retreat. (C) A repulsive force pushes the two categories apart, with greatest strength in the area of overlap between the categories, allowing the near side of the Pushee to retreat. In the case of single-category movement, there is only one category, and (C) does not apply. (D) A squeezing force ensures that the categories do not become too wide or skewed. We propose that word frequency-based asymmetries in regular sound change can be generated by allowing the repulsive force, which we relate to robustness of spoken word recognition, to vary with frequency. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

balance an intrusive force, a spreading force, and a squeezing force. A model of a two-category system must initiate and balance these three forces, plus a fourth, repulsive force. Since the single-category case is subsumed under the two-category case, we only discuss the two-category case henceforth.

The intrusive force causes the Pusher to move toward the Pushee, initiating the interaction. On its own, the intrusive force will not trigger the Pushee to move. The spreading force helps to address this issue by spreading each category outward, causing the outer edge of the Pushee to move away. However, this still will not allow the Pushee as a whole to move; instead, the distance between categories will decrease and their overlap will grow. The repulsive force solves these problems by pushing the categories away from one another, allowing both categories to move while remaining the same distance apart. But the repulsive force will also increase category skewness, since it acts more on the part of each category in the overlapping region than on the other part. The squeezing force counters this increase in skewness by squeezing each

category inward, with the long tail (the tail outside of the overlapping region) being squeezed more than the short tail (the tail in the overlapping region).[7] The squeezing force also acts with the repulsive force to ensure that the categories do not become too wide under the spreading force.

The repulsive force pushes overlapping categories away from each other, enforcing an aversion to acoustic ambiguity. Our main proposal in this paper will rest on the notion that high-frequency words experience this aversion less than low-frequency words, as they can be more robustly recognized in the face of acoustic ambiguity than low-frequency words (see Section 5.3). In other words, high-frequency words are less sensitive to the repulsive force than low-frequency words. In Section 5.5, we test whether incorporating such an asymmetry in a two-category model that generates and successfully balances the four forces in Fig. 2 causes high-frequency words to change faster than low-frequency words in the Pusher and vice versa in the Pushee.

2.5. Model implementation: Exemplar Theory

We implement the high-level forces discussed above using representations based on Exemplar Theory, which proposes that categorization of a perceived stimulus entails the comparison with *exemplars* – episodic traces of experienced instances – of other stimuli, stored in memory (e.g. Nosofsky, 1986). For speech perception, this is taken to mean that listeners store richly-detailed memories of spoken words as they experience them, which they use as a basis of comparison for categorizing other instances of spoken words (Goldinger, 1996; Johnson, 1997). Pierrehumbert (2001, 2002) proposed that speech production could also be exemplar-based, with the production of a spoken word drawing on the same exemplars that would be used in the perception of it. The joining of production and perception in this way creates a closed loop, within which the processes acting on individual exemplars aggregate to yield forces acting across entire distributions of words containing the same phonological segment. The modeling of these forces as emergent over atomic exemplars in the distribution provides a convenient way to capture word-specific patterns of sound change whilst also capturing larger patterns that recur across words (Pierrehumbert, 2002). Exemplar-based models of the perception-production loop have been applied to various kinds of sound change (Ettlinger, 2007; Harrington et al., 2018; Pierrehumbert, 2001, 2002; Sóskuthy, 2013; Tupper, 2015; Wedel, 2004, 2006, 2012; Wedel & Fatkullin, 2017).

A common criticism of exemplar-based models is that they privilege high-frequency words, because their higher rate of occurrence leads them to be represented more densely in memory. By consequence, exemplar-based models are often interpreted to predict that high-frequency words should change fastest in every regular sound change, which draws criticism for predicting frequency effects where there aren't any (Abramowicz, 2007; Bermúdez-Otero et al., 2015; Dinkin, 2008; Tamminga, 2014). In showing that our model generates different effects of word frequency in different kinds of sound change, we aim to dispel the misunderstandings underlying this criticism and point out the value that exemplar-based computational modeling can bring to the study of sound change.

3. Model description

Our model is formulated as a production-perception loop (Pierrehumbert, 2001) and consists of a cycle of processes applying to exemplars (one per iteration). In this section, we describe at a high level the representations in the model, the processes constituting production

and perception, and the way in which these processes yield the forces that drive the evolution of the system.

3.1. Representations in the model

The model describes how the realization of a phoneme occurring within words changes over time as those words are used in successful communication. For concrete illustration, we assume that the phoneme in question is a vowel, and the words in question are monosyllabic. The model contains three levels of representation in memory: *category*, *type*, and *exemplar*. The number of exemplars of a particular type reflects the *type frequency*, and the exemplars are arranged within a perceptual-acoustic *exemplar space*. We include a glossary of representation terms in Table 1, along with measures of the representations in simulations, for ease of reference.

3.1.1. Categories

A *category* corresponds to a phoneme, representing an abstract generalization over experienced instances of that phoneme (in words). For example, the category representation of the phoneme /æ/ is a generalization over experiences with words like *map*, *lab*, *cat*, etc. Our simulations involve one or two categories; in the two-category case, a Pusher category moves toward a Pushee category. For discussion of systems with more than two categories, see Appendix A.1.

3.1.2. Types

A category consists of a set of *types*, which represent abstract templates for words containing the corresponding phoneme. There are two components to a type: the phonological frame, which specifies the parts of the word that are *not* at issue, and the category to which the type belongs, which specifies the phoneme that *is* at issue. For example, the type corresponding to the word *map* has the phonological frame /m_p/ and belongs to the category /æ/. For simplicity, we assume that there are no minimal pairs, so that the category membership of a given type may be uniquely determined from its phonological frame (see Appendix A.2 for discussion).

Our simulations include 92 types per category, approximately one-tenth of the number in the New Zealand vowel shift data[8]; see Section S1.1 of the Supplementary Materials for details.

3.1.3. Exemplars

Each type is substantiated by *exemplars*, which are detailed memory traces of instances of that type. Exemplars encode experienced perceptual-acoustic realizations of the category phoneme. For example, the representation for the word *map* includes a number of remembered instances of spoken "map", each one containing a slightly different realization of the vowel /æ/. Our simulations include 492 exemplars per category (see Section S1.1 of the Supplementary Materials for details).

3.1.4. Frequency

Different types have different numbers of exemplars; we follow the multiple-trace hypothesis (Hintzman & Block, 1971) in assuming that the number of exemplars for a given type represents that type's frequency. For the simulations presented in this paper, we modeled the number of exemplars corresponding to a word on the log-frequency of that word in a large corpus, yielding frequencies ranging from 1 to 12 (see Section S1.1 of the Supplementary Materials). We use the same type frequencies for perception and production (see Appendix A.3 for discussion).

---

[7] In our model, the squeezing force squeezes toward the mode. In the Section S1.2.8 of the Supplementary Materials, we discuss why this is preferable to the approach taken by previous models, where it squeezes toward the mean.

[8] Increasing the number of types does not change the qualitative results of the model, but slows down its evolution by an approximately proportionate amount; see Sections S3.2 and S5.2.3 of the Supplementary Materials for illustrative simulation and mathematical discussion, respectively.

Table 1
Glossary of representation terms.

Term	Meaning	Example	Simulations
Category	A generalization over experienced instances of a phoneme (e.g. a vowel), stored in memory	/æ/	1–2 categories
Type	An abstract template for a word containing a particular phoneme, collecting together experienced instances of that word in memory. Contains information about the *frame* (e.g. onset and coda consonants of a monosyllabic word) and the category (e.g. nucleus vowel)	*map*	92/category
Exemplar	A memory trace of an experienced instance of a particular type, i.e. a spoken word	"map"	492/category
Type frequency	The number of exemplars of a given type. Based on word log-frequency in a large corpus (see Section S1.1 of the Supplementary Materials)		Range: 1–12
Exemplar space	The distribution of exemplars across a granularized perceptual-acoustic dimension (e.g. vowel F1). Assumed to be shared across perception and production		Grain: 0.1

3.1.5. Exemplar space

The model tracks the distribution of exemplars in a perceptual-acoustic *exemplar space*. Following Kruschke (1992), the exemplar space is granularized, such that acoustic values that are different but nevertheless perceived identically are all represented by a single, shared value (see Pierrehumbert, 2001). For this early work, we make the simplifying assumption that the exemplar space is one-dimensional, e.g. corresponding to vowel F1; for discussion, see Appendix A.1.

We assume a single exemplar space, shared across production and perception. This can be interpreted as a single agent talking to herself, or as an aggregate over a homogeneous community talking amongst itself. The modeling of multiple agents with distinct spaces is left for future work.

The model is initialized with a distribution of exemplars for each category, constructed in such a way as to avoid asymmetries across categories, across frequency classes, or across exemplars within a frequency class (see Section S1.1 of the Supplementary Materials). The initial width of the categories, σ, and the initial distance between them, μ, are parameters of the model.

3.2. Processes in the model

Each iteration of the model begins with the production of a *token* – an instance of a type with a particular *target* acoustic value – based on the existing exemplar space. The produced token is then transmitted to the listener, as the acoustic value of an unknown phoneme residing in a known phonological frame.[9] In perception, the listener uses the value and frame to recover the type (and thus the category) intended by the speaker, and then decides whether the token should be stored as an exemplar of that type and thus update the corresponding category distribution.

Together, production and perception form a closed loop, gradually updating the distribution of exemplars in the space through the generation and storage of tokens. This loop is composed of multiple processes, both on the production side and the perception side, as illustrated in Fig. 3. It is these processes that yield the category-level forces, as illustrated in Fig. 4.

We describe the processes in the model in the following subsections. For technical details, see Section S1.2 of the Supplementary Materials.

3.2.1. Type and target selection

The speaker first selects a type (weighted by frequency), and then selects an exemplar of that type to constitute the initial production target. Both selections are random and independent of previous selections, together constituting a random draw from the set of exemplars.

Prior to realization, the target value is adjusted under two influences, bias and imprecision.

3.2.2. Bias

The first adjustment of the target value is due to *bias*, which shifts the target by a small and consistent amount. Bias represents external influences such as reduction of articulatory effort. At the level of the category distribution, it yields the intrusive force (henceforth, the *bias force*).

We apply bias to all[10] productions in the single-category case, and to all productions of the Pusher category in the two-category case. The size of the bias, β, is a parameter of the model.

3.2.3. Imprecision

The second adjustment of the target value is due to *imprecision*, which shifts the target by a small (random) amount in either direction. Imprecision represents natural variability in the application of motor routines in realization.[11] At the level of the category distribution, it yields the spreading force (henceforth, the *imprecision force*).

We implement imprecision through the addition of random noise to the target, for all productions. The degree of imprecision, ι, is a parameter of the model.

3.2.4. Activation

For the listener, the incoming token *activates* exemplars of both categories within a window around the target, with exemplars near the target activated more than those far away. These activations are aggregated within each category to yield overall category activation, which underlies key processes in perception. The size of the activation window, α, is a parameter of the model.

3.2.5. Identification

Based on the phonological frame, the listener *identifies* the type[12] corresponding to the token. Not every identified token is stored as an exemplar, updating the category representation; for that, the token must be "good" enough, i.e. must pass the discriminability and typicality evaluations.

---

[9] For simplicity, we assume that both the acoustic value and the phonological frame of the token are perceived exactly as produced.

[10] A reviewer asks whether applying bias selectively to a subset of types would generate category movement. We believe so, provided that the subset of biased types is sufficiently large and the bias force sufficiently small (relative to the typicality force, which promotes category cohesiveness). When these conditions are not met – such as in phonologically-conditioned changes or isolated socially-meaningful realizations – we expect the category to split. The investigation of such cases is beyond the scope of the present paper and is left for future work.

[11] In principle, imprecision may be experienced by the speaker or by the listener. The number of imprecision processes does not affect the high-level model behavior, so we choose a single process for simplicity. We choose to locate this process in the speaker, to make it clear that our claim of the listener's centrality to sound change is based on processes that have strong justification for being listener-based.

[12] In the present model, there is just one candidate type for the token – which is the type intended by the speaker – because phonological frames are perfectly transmitted and there are no minimal pairs. In principle, however, a set of types may be plausible candidates, and each may be assessed (in subsequent evaluations) for the extent to which it is compatible with the transmitted token (Norris & McQueen, 2008). We discuss the introduction of architecture to handle multiple candidate types in Section S3.4.1 of the Supplementary Materials.
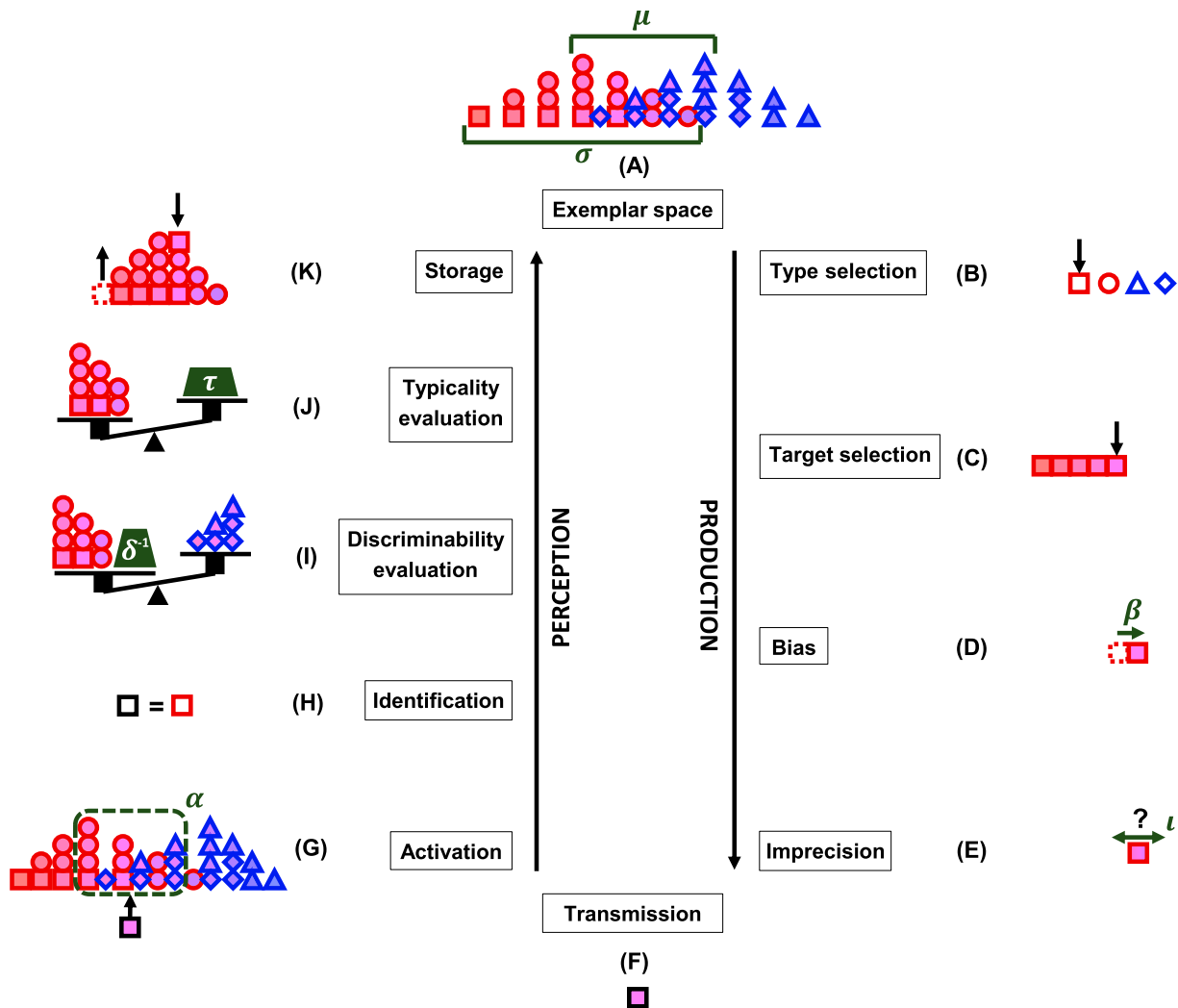
Fig. 3. Schematic illustration of processes in the model, forming a closed loop between production and perception. Outline colors represent phoneme category membership (e.g. /æ/), shapes represent phonological frame (e.g. /m_p/) – so that colored shape-outlines represent types (e.g. "map") – and fill colors and horizontal positions represent perceptual-acoustic value (e.g. vowel F1). Dark green components with Greek letters indicate parameters of the model. (A) Two partially-overlapping categories exist in an exemplar space. The initial category width, σ, and distance between categories, μ, are parameters. (B) The speaker randomly selects a type for production, according to its frequency. (C) An exemplar of that type is randomly selected to provide an acoustic target for the production. (D) For the Pusher, the target is shifted by a constant bias toward the Pushee. The size of the bias, β, is a parameter. (E) The actual realization of the target is imprecise, causing it to shift by a random amount in either direction. The degree of imprecision, ι, is a parameter. (F) The realized token is transmitted to the listener, with its acoustic value and phonological frame but without its category membership. (G) The listener locates the token in their exemplar space, activating surrounding exemplars of both categories within a fixed activation window. The size of the activation window, α, is a parameter. (H) The candidate type of the token is identified based on context (represented here by phonological frame), yielding identification of the intended category. (I) The activation of the intended category (red) is compared to the activation of the other category (blue); if the ratio of activations is below a fixed discriminability threshold, the token is unlikely to be stored. The discriminability threshold, δ, is a parameter. (J) The activation of the intended category is compared to a typicality threshold; if the activation is below the threshold, the token is unlikely to be stored. The typicality threshold, τ, is a parameter. (K) If the token is sufficiently discriminable and typical, it is stored in the listener's exemplar space, replacing a random exemplar of the same type. In our model, the speaker and listener are the same, and thus storage updates the exemplar space for future production and perception. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2.6. Discriminability evaluation

The discriminability evaluation poses the question: how likely is the token to be a realization of its identified category, as opposed to the other category, based on its acoustic value? It follows results in speech perception that tokens that are acoustically ambiguous between categories incur processing costs, causing errors and delays in recognition (e.g. Connine, Blasko, & Hall, 1991; see also Section 5.3 for interactions with word frequency). Tokens that do not pass the evaluation are not stored, and hence do not update the category distribution.

The discriminability evaluation is probabilistic, based on the ratio of category activations (identified/other). Hence, tokens outside of the region of category overlap (where the ratio is large) are more likely to pass than tokens inside it (where the ratio is small). At the level of the category distribution, this asymmetry yields the repulsive force (henceforth, the *discriminability force*).

The evaluation proceeds by comparing the category activation ratio to a discriminability threshold, δ, which is a parameter of the model. The size of δ determines the size of the discriminability force: as δ grows higher, passing the evaluation becomes harder, and the force grows stronger.[13]

[13] The net discriminability force also grows with the size of the activation window, α, as a wider window encapsulates more exemplars and yields a category activation ratio closer to 1, which is lower for most tokens.
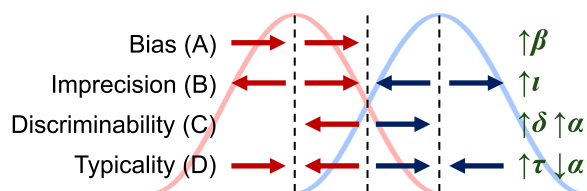
Fig. 4. Iterated over time, processes in the model exert forces on the distributions of exemplars within each category. (A) Bias pushes one category toward the other. This force increases with β. (B) Imprecision spreads each category outward. This force increases with ι. (C) Discriminability pushes exemplars out of the region of overlap between categories, repelling categories away from one another. This force increases with δ and α. (D) Typicality lightens the tails of distributions, squeezing each category toward its mode and countering skewness. This force increases with τ and decreases with α.

3.2.7. Typicality evaluation

The typicality evaluation poses the question: how good is the token as a realization of its identified category, in absolute terms? It follows results in speech perception that tokens that are "good" instances of their category – i.e. that are similar to many other experienced instances of the category – are encoded strongly in memory (e.g. Clopper, Tamati, & Pierrehumbert, 2016), giving them advantages in immediate processing (e.g. Johnson, 2006) and long-term recall (e.g. Sumner, Kim, King, & McGowan, 2014). Tokens that are poor instances of their category, i.e. that do not pass the evaluation, are not stored.

The typicality evaluation is probabilistic, based on the activation of the identified category. Hence, tokens that are near the mode of the category (where activation is high) are more likely to pass than tokens that are far from it (where activation is low). At the level of the category distribution, this asymmetry yields the squeezing force (henceforth, the *typicality force*).

The evaluation proceeds by comparing (normalized) category activation to a typicality threshold, τ, which is a parameter of the model. The size of τ determines the size of the typicality force: as τ grows higher, passing the evaluation becomes harder, and the force grows stronger.[14]

3.2.8. Storage

The token updates existing category representations by being *stored* as an exemplar of the identified category, but only if it passes both the discriminability and typicality evaluations. It follows that "poor" (indiscriminable and/or atypical) productions have much less influence on representations – and thus are much less likely to be repeated – than "good" productions.

When a token is stored, it overwrites a random exemplar of the same type (see e.g. Landauer, 1986, for discussion of overwriting as a principle of memory). While this approach is a clear simplification, it yields few differences from alternatives; for discussion, see Appendix A.4.

3.3. Comparison to existing models

At a high level, most of the representations and processes in our model are common to previously-proposed exemplar dynamics models (Ettlinger, 2007; Pierrehumbert, 2001, 2002; Sóskuthy, 2013; Tupper, 2015; Wedel, 2004, 2006, 2012; Wedel & Fatkullin, 2017), though some low-level details of implementation differ (for fine-grained comparison, see Section S1.2 of the Supplementary Materials). There are two components of our model that stand out from this commonality.

Firstly, our model includes both category- and type-level

representations, enabling us to treat word frequency separately from phoneme frequency. With a single exception (Sóskuthy, 2014), previous models have (explicitly or implicitly) treated types as categories and thus confounded two potential sources of frequency effects. This lack of distinction contributed to the claim by Pierrehumbert (2001) that high-frequency words change fastest under production bias in regular sound change; for discussion, see Section S5.2.5 of the Supplementary Materials.

Secondly, our model generates the squeezing force with the novel process of typicality evaluation, rather than the standard process of entrenchment[15]. Our process better facilitates the maintenance of category overlap and shape (see Section S1.2.8 of the Supplementary Materials), which previous models have struggled with; for example, the model presented by Tupper (2015) generates skewed, barely-overlapping categories, and Pierrehumbert (2002, p. 133) states that she "[has] not actually been able to find a parameter range for this model which shows stable overlapping distributions".

4. Modeling single category movement

We begin by modeling the most basic case: movement of a single category. Recall from Section 2.3 that the basic desiderata for such a model are that it: (i) generates movement of the category; and (ii) maintains the shape (width & skewness) of the category. In this section, we describe our approach to meeting these desiderata, and we show that the model captures the key frequency-related property of Manchester /t/-glottaling – change at the same rate across words of all frequencies.

4.1. Approach

As established in Section 3, the overall evolution of a system involving a single biased category is defined by three forces stemming from production and perception processes, whose strengths are determined by the value of key model parameters (β, ι, α, and τ). The task of meeting the desiderata for a model of single category movement therefore reduces to the task of finding parameter values that balance these forces against one another, generating consistent non-distorting category movement. To accomplish this task, we tuned sets of parameters in a stepwise process; for detailed discussion, see Sections S2.1–2.2 of the Supplementary Materials.

4.2. Results and discussion

The tuning process identified numerous parameter values yielding behavior that met our desiderata. Taking the fact that the model generates appropriate single category movement as sufficient foundation, we now turn to our primary interest: the effect of word frequency on rate of change. We hypothesized that placing emphasis on the listener would predict no effect, since the movement has no adverse implications for the listener (i.e. yields no changes in the discriminability of tokens).

Fig. 5 shows how both low- and high-frequency types move over time, averaged across 1000 simulations, for representative parameter settings under different degrees of bias. The centroids of the low- and high-frequency sub-distributions change in parallel (i.e. at the same rate), supporting our hypothesis that sound change without adverse implications for the listener is frequency-independent. We take this result to mean that a listener-based focus can explain the lack of word

---

[14] The typicality force also grows as the size of the activation window, α, shrinks. A narrower window encapsulates fewer exemplars and yields lower category activation.

[15] Entrenchment is a form of averaging, such as might occur due to practice effects in production (Pierrehumbert, 2001), and generates the squeezing force in most exemplar dynamics models. We exclude entrenchment for present purposes, but have included it in the accompanying code; see Section S4 of the Supplementary Materials for discussion.
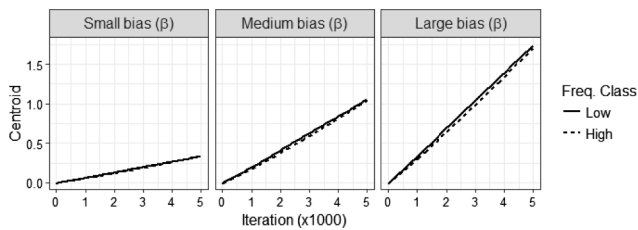
Fig. 5. Results of simulations with a single category ($\sigma = 0.8$) subject to varying degrees of bias, illustrating differences between low-frequency (solid) and high-frequency (dashed) types. For all degrees of bias, the centroid of the category distribution advances at the same rate for both low- and high-frequency types.

frequency effect on rates[16] of /t/-glottaling in Manchester English (Bermúdez-Otero et al., 2015).

Our result is critically different to that predicted by prominent usage-based models and theories (Bybee, 2002; Phillips, 1984; Pierrehumbert, 2001), in which the higher rate of production of high-frequency words translates into a higher rate of response to bias. The primary reason for this difference is that we model both the type (word) level and the category (phoneme) level, allowing types of different frequencies to be represented by different numbers of exemplars. Even though high-frequency types are produced (with bias) more often than low-frequency types, they are also represented by more exemplars (Hintzman & Block, 1971). Thus, an isolated production has less influence on the representation of a high-frequency type than on that of a low-frequency type, counterbalancing the difference in rates of production (see also Sóskuthy, 2014).[17] We illustrate the differences between previous usage-based models and our model schematically in Fig. 6; for discussion from a mathematical point of view, see Section S5.2 of the Supplementary Materials.

5. Modeling two-category interaction

We now turn to the modeling of two-category interaction. Recall from Section 2.3 that the basic desiderata for such a model are that it: (i) generates movement of one category in response to the other; (ii) maintains the distance between the two categories; (iii) maintains the shape (width and skewness) of the categories; and (iv) maintains the overlap of the categories. To our knowledge, no other exemplar-based model reported in the literature has met all of these desiderata, due primarily to difficulties with maintaining overlap of non-skewed categories.

In this section, we show that our model can meet the desiderata. We argue that perceptual processes are central to this success, underscoring the importance of the listener to sound change. We then show that enriching the basic model with an empirically-grounded perceptual asymmetry allows for the generation of word frequency effects on rate of change that resemble those seen in New Zealand English /t/-tapping (Hay & Foulkes, 2016) and /ɛ/-raising (Hay et al., 2015).

---

[16] We emphasize that our focus is on the lack of word frequency effects on *rate of change*, and not on the existence of a stable frequency effect whereby high-frequency words exhibit a fixed amount more /t/-glottaling than low-frequency words at every point in time. To account for this stable effect, the model would have to assume that high-frequency words are more prone to hypoarticulation than low-frequency words (e.g. Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Gahl, 2008) and that the initial conditions reflect this asymmetry (rather than being neutral, as at present). The same assumptions are required in the explanation put forward by Bermúdez-Otero et al. (2015).

[17] While the centroids of the sub-distributions of exemplars of high- and low-frequency types changed at the same rate in our simulations, the sub-distributions themselves did not evolve identically, due to a side-effect of our assumptions about production and storage. For details and discussion, see Appendix B.
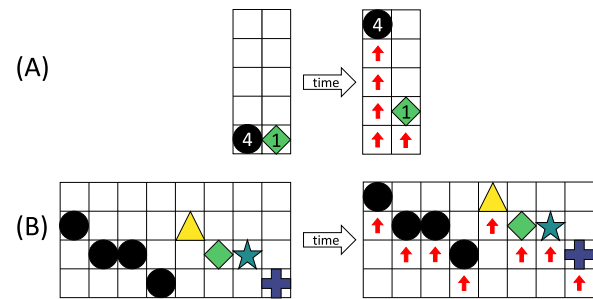

Fig. 6. Schematic illustration of the intuition underlying the interaction of type frequency and bias in prominent usage-based models (A; e.g. Pierrehumbert, 2001) and the present model (B). Different types are represented by different shapes and colors, acoustic value is represented by vertical position, and movement due to bias is represented by red arrows (one per production of the type). Each case depicts a high-frequency type (black circles; frequency 4) and some low-frequency types (colored angled shapes; frequency 1). In (A), shapes represent the location of the mean acoustic value for each type, and type frequency is indicated by numbers; in (B), the exemplar distribution of each type is represented by individual exemplars, and type frequency is indicated by number of exemplars. The left panel shows an initial condition, and the right panel shows the expected (average) result after the amount of time it takes to produce the high-frequency type 4 times. (A) Intuition underlying prominent models; comparing a single type of each frequency. In the expected amount of time it takes to produce the high-frequency type 4 times (5 iterations), the low-frequency type is expected to be produced once. Each production of a type adds bias and updates that type's mean. Since the high-frequency type is produced 4 times more often than the low-frequency type, it is subjected to 4 times as much bias. The high-frequency type thus evolves at a faster rate than the low-frequency type. (B) Present model; comparing the aggregate over exemplars of each frequency. In the expected amount of time it takes to produce the high-frequency type 4 times (8 iterations), the four low-frequency types are each expected to be produced once. Each production of a type adds bias to a single exemplar of that type. Since each exemplar of each type is produced once, the bias is distributed over the exemplars. While the high-frequency type is subjected to more bias than any low-frequency type, it distributes this bias over 4 exemplars, which is equivalent to the distribution of bias over 4 exemplars of the 4 different low-frequency types in aggregate. The sub-distributions of high- and low-frequency exemplars thus evolve at the same rate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.1. Basic model: approach

The general approach for modeling two-category interactions is similar to that used for modeling single-category movement (Section 4.1). The task is to find values of key model parameters (β, ι, α, δ, and τ) that balance the forces acting on the two category distributions. As in the single-category case, we accomplished this task by tuning sets of parameters in a stepwise process, ignoring potential frequency effects; for detailed discussion, see Sections S2.3–2.4 of the Supplementary Materials.

5.2. Basic model: results and discussion

The tuning process identified numerous parameter values that yielded interlinked category movement with maintenance of category width, shape, distance, and overlap; to our knowledge, a first in the exemplar dynamics literature. An example is illustrated in Fig. 7.

We attribute the success of our model in meeting the desiderata to three aspects of the perceptual processes undertaken by the listener. Firstly, by using a discriminability threshold $\delta < 1$, our model induces a lexical bias (Ganong, 1980) in cases of low discriminability, effectively shrinking the part of the overlapping region between categories that creates instability for the perception of attested (real-word) types. Secondly, by not storing tokens that fail the discriminability evaluation (i.e. tokens that are likely to be recognized as nonwords), our model
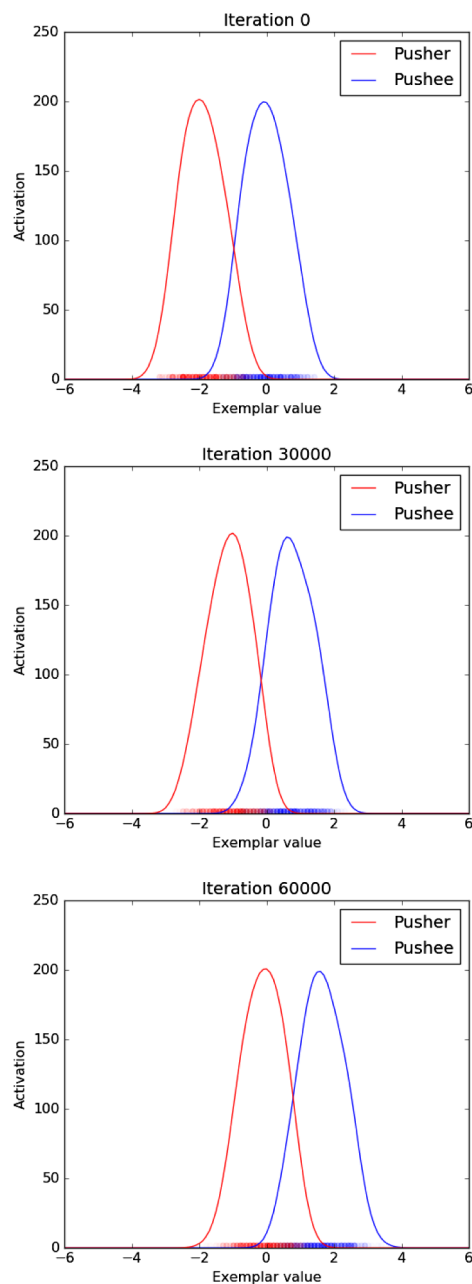
Fig. 7. The evolution of exemplar distributions (rugs on horizontal axis) and corresponding activation fields for one run of the model (parameter set (2) from Table S3 of the Supplementary Materials). Over time (from top panel to bottom), the two categories move to the right, maintaining their distance from one another, their degree of overlap, and their widths and skewnesses.

avoids skewness-inducing overpopulation of the overlapping region between categories (see e.g. results of "competition with discards" in Tupper, 2015). Thirdly, by including the novel process of typicality evaluation, our model generates a squeezing force that keeps skewness in check while facilitating overlap (see Section S1.2.8 of the Supplementary Materials for comparison to the standard alternative). The importance of these perceptual processes supports our claim that the listener plays a central role in sound change.

Further support for the centrality of the listener is provided by the fact that, without the listener, categories would not interact at all in our model: the Pusher would simply float over the Pushee, because the speaker samples production targets without concern for potential ambiguity. The listener prevents this behavior by providing an indirect,

non-teleological influence of perceptual filtering on production: whenever the speaker is ambiguous, the listener is unlikely to store the token, and is thus unlikely to use it as a basis for future productions. The listener thus drives category interaction in our model by creating category repulsion via the discriminability force, with the speaker's constant bias serving to ensure that interaction persists in the face of this repulsion.[18] More generally, the listener drives any self-organizational response to the system as a whole, as perceptual processes involve the activation of multiple exemplars (from both categories), whereas production processes involve no more than the single initial target exemplar.

In the absence of additional mechanisms, the basic model generates frequency effects on rate of change that are the reverse of those seen empirically: high-frequency types change slower than low-frequency types in the Pusher and faster in the Pushee (Fig. 8). These strength of these effects is due to our assumptions about production and storage; see Appendix B for discussion.

5.3. Enhanced model: motivation

To successfully model the frequency effects from New Zealand English /t/-tapping (Hay & Foulkes, 2016) and /ɛ/-raising (Hay et al., 2015) respectively, high-frequency types would have to change faster than low-frequency types in the Pusher and slower than low-frequency types in the Pushee. The basic model generated the reverse effects. We argue that the basic model failed because it does not incorporate empirically supported frequency-based asymmetries in perception.

The literature contains numerous empirical results showing that high-frequency words are privileged over low-frequency words in speech perception, both when there is no salient lexical competitor, and when there is. In situations without a salient lexical competitor, relative to low-frequency words, high-frequency words are intelligible in larger amounts of masking noise (Howes, 1957) and are classified as real words more often (Luce & Pisoni, 1998) and faster (Forster & Chambers, 1973) in lexical decision. In situations where multiple salient words compete for recognition, higher-frequency words attract more attention early in processing (Dahan, Magnuson, & Tanenhaus, 2001) and are favored responses to degraded stimuli (Savin, 1963) or stimuli from a dialect other than one's own (Clopper, Pierrehumbert, & Tamati, 2010).

Furthermore, a series of phonetic categorization studies have shown word frequency effects in the mapping of an acoustically ambiguous stimulus to one of two words in a minimal pair. Fox (1984) observed that, when presented with ambiguous stimuli on a "bad"-"dad" continuum, listeners were more biased toward *bad* responses than expected (based on their responses to a /bæ/-/dæ/ continuum). He suggested this might be because *bad* is more frequent than *dad*. Connine et al. (1993) provided support for this suggestion from a range of continua between high- and low-frequency words that differ in initial stop voicing (e.g. "best"-"pest"). Ambiguous stimuli on these continua were more likely to trigger the high-frequency word response (e.g. *best*) than the low-frequency word response (e.g. *pest*). Similar results were found by VanDam (2007). Finally, de Marneffe, Tomlinson, Tice, and Sumner (2011) replicated this result using manipulated French-accented English words with final stops (e.g. "tag" and "tack"), showing that the high-frequency response bias is not limited to situations where the stimulus begins with an ambiguous sound. Furthermore, they showed

---

[18] A reviewer asks what would happen in a system without bias. Without bias, mutual category repulsion stemming from the discriminability force would cause category separation (see Section S3.3.2 of the Supplementary Materials). With limits on the space of possible articulations, the system would eventually stabilize to maximally distribute the categories, as in Dispersion Theory (Liljencrants & Lindblom, 1972). Thus, removing bias limits category interaction to a "default" basis, removing the possibility for the sustained and targeted movement that we model here.
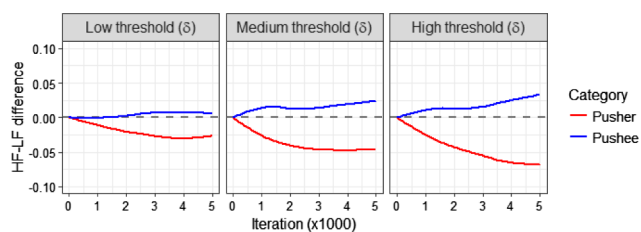
Fig. 8. Representative results of simulations in the basic model with two categories for different discriminability thresholds (parameter sets (4), (10), and (16) from Table S3 of the Supplementary Materials), illustrating the extent to which high-frequency types are ahead of low-frequency types in the Pusher (red) or Pushee (blue). A positive slope represents a faster rate of change of high-frequency types compared to low-frequency types. For all discriminability thresholds, high-frequency types change slower than low-frequency types in the Pusher and faster than low-frequency types in the Pushee – the reverse of the frequency effects observed empirically. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that the bias is not limited to minimal pairs with extreme frequency differences, but is found across minimal pairs, with strength related to the ratio of word frequencies.

We take this set of results to imply that – all else being equal, i.e. absent effects of speech style or semantic context – the perceptual system is biased toward the recognition of high-frequency words, especially in the case of acoustically ambiguous tokens. In the following sections, we show that enriching the basic model to encode such a bias allows it to generate the frequency effects on rate of change seen empirically in /t/-tapping and /ɛ/-raising in New Zealand English.

5.4. Enhanced model: approach

We assume that acoustically ambiguous tokens of high-frequency words are more robustly recognized and stored than similarly-ambiguous tokens of low-frequency words (see also Hay et al., 2015, for more discussion). We encode this perceptual asymmetry in our model by varying the discriminability threshold, δ, with type frequency.[19] Specifically, we give tokens of high-frequency types lower δ than tokens of low-frequency types, making them more *discriminable*, i.e. more likely to pass the discriminability evaluation and be stored when encountered in the overlapping region between categories. This assumption has no implications for the case of single-category movement (Section 4), as the discriminability evaluation cannot fail in a system containing only one category.

To test the hypothesis that frequency-based asymmetries in discriminability could give rise to empirically observed frequency effects on rate of change, we conducted simulations with frequency-sensitive δ, keeping all other parameters fixed at the previously-tuned values from Section 5.1. We constructed 15 frequency-sensitive δ functions, which are illustrated in Fig. 9A and described fully in Section S1.3 of the Supplementary Materials.

5.5. Enhanced model: results and discussion

The results of varying δ with type frequency are shown in Fig. 9B. When average discriminability is sufficiently high and high-frequency types are sufficiently more discriminable than low-frequency types, the model generates robust frequency effects resembling those seen

empirically. In Section S3.1 of the Supplementary Materials, we show how these effects can be reinforced by further empirically-supported asymmetries in the typicality evaluation. The general result is that, given appropriate frequency-based perceptual asymmetries, high-frequency types change faster than low-frequency types in the Pusher and slower than low-frequency types in the Pushee.

The result obtains because, though each category distribution as a whole is subjected to a given discriminability force, types of different frequencies are differentially sensitive to it. High-frequency types are less sensitive to the discriminability force and low-frequency types are more sensitive to it. To balance category-level forces, the sub-distribution of high-frequency exemplars is shifted closer to the overlapping region, where the local discriminability force is larger, and vice versa for the sub-distribution of low-frequency exemplars.[20] The size of the frequency effect on rate of change is thus a function both of the average discriminability, which determines the size of the discriminability force on average, and of the degree of frequency-based discriminability asymmetry, which determines the difference in sensitivity to the discriminability force.

As in Section 5.2, we emphasize that the results here are driven by the listener, not the speaker.[21] In our model, high-frequency types and low-frequency types all show the same sensitivity to speaker-based forces (bias and imprecision), and there is no impetus for the speaker to produce tokens of high-frequency types in a hypoarticulated manner or tokens of low-frequency types in a hyperarticulated manner (cf. Lindblom, 1990). Rather, tokens of high-frequency types are more robustly recognized than tokens of low-frequency types, leading them to be more likely to be stored when they are in the overlapping region between categories. Consequently, the high-frequency sub-distribution will come to be dominated less by "clear" exemplars (from outside of the overlapping region) than the low-frequency sub-distribution, and the asymmetry in perception will drive asymmetries in production without the speaker ever having an intention to adjust the clarity of her productions.
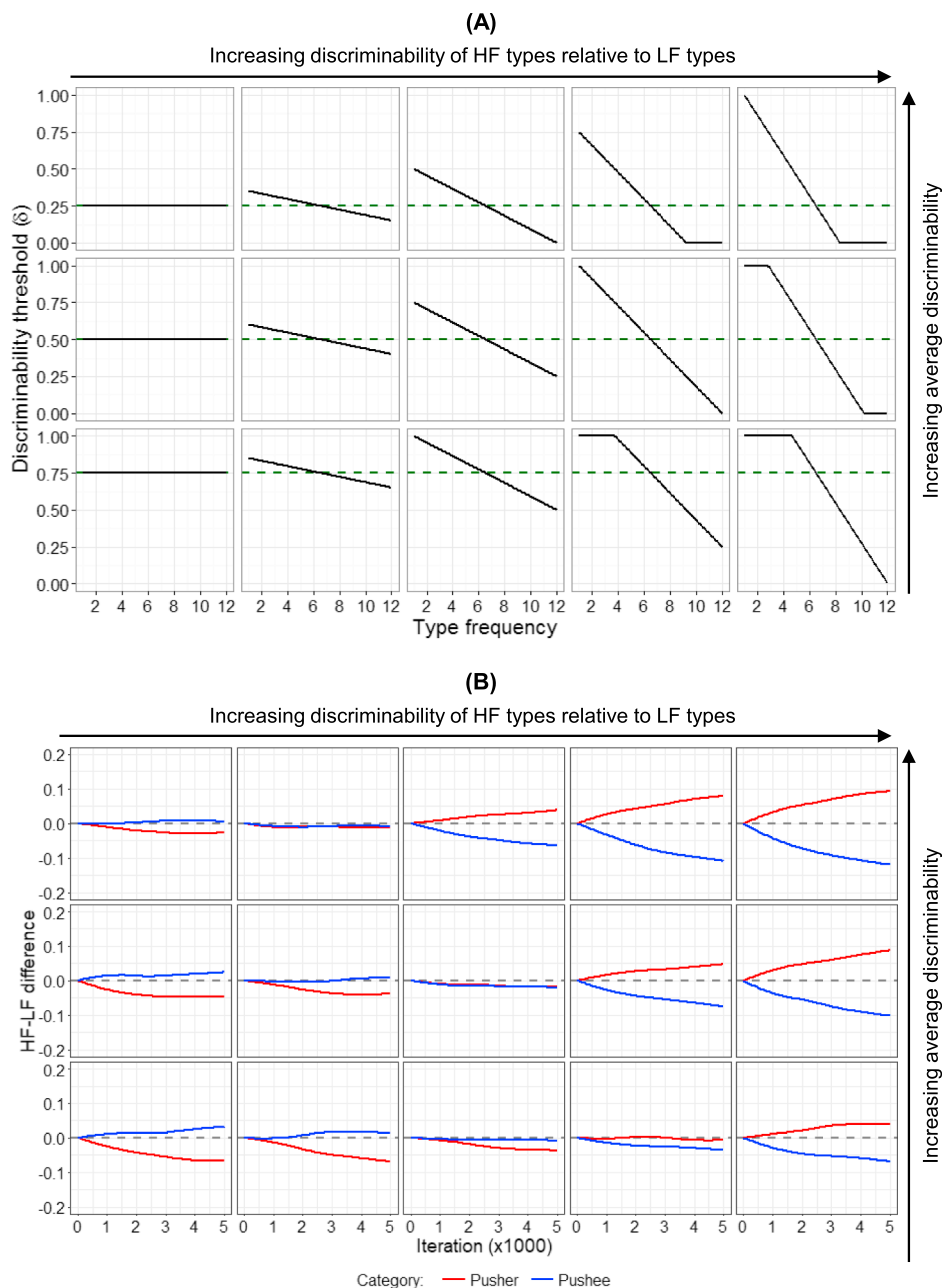
6. General discussion

We set out to explore whether patterns of regular sound change could be driven by processes within the listener, rather than by processes within the speaker. In particular, we asked whether an emphasis on the listener could explain why high-frequency words change at the same rate as low-frequency words in some kinds of regular sound change, faster than low-frequency words in other kinds of regular sound change, and slower than low-frequency words in yet other kinds of change. To address this question, we constructed an exemplar-based computational model of single-category movement and two-category interactions with empirically-motivated high-level properties, into which we incorporated a bias for high-frequency words to be more robustly recognized than low-frequency words in the face of acoustic ambiguity. We showed that, with such a bias in the listener, the model successfully explained the different effects of word frequency on rate of change in /t/-glottaling in Manchester English (Bermúdez-Otero et al., 2015), /t/-tapping in New Zealand English (Hay & Foulkes, 2016), and /ɛ/-raising in New Zealand English (Hay et al., 2015).

In this section, we discuss the model's predictions for word

---

[19] The high-level assumption that δ varies with type frequency is sufficient to show that asymmetries in discriminability yield asymmetries in rates of change. Ultimately, however, we would like these asymmetries to emerge mechanistically from the exemplar-based architecture (i.e. with a fixed δ for all types). In Section S1.3.1 of the Supplementary Materials, we discuss two mechanisms by which such asymmetries could emerge.

[20] A reviewer asks if this separation of exemplar sub-distributions could lead to a category split. The answer is no, because all exemplars within a category are subjected to the same typicality force, which squeezes them toward a single mode. We also emphasize that the model treats frequency as a continuous variable, so that a category is made up of sub-distributions spanning the entire frequency range, not just the two extremes that we use for illustration.

[21] We provide further evidence that frequency effects on rate of change are driven by the listener, not the speaker, with additional simulations in Section S3.3 of the Supplementary Materials. We observe the same frequency effects both in the absence of category movement and in the absence of speaker bias.

(A)

Increasing discriminability of HF types relative to LF types

(B)

Increasing discriminability of HF types relative to LF types

Category: Pusher   Pushee

Fig. 9. Details and results of treating discriminability threshold (δ) as a function of type frequency. (A) δ functions investigated (black lines). Lower δ indicates greater discriminability. Across all panels in a given row, δ is kept constant for median-frequency types (dashed green lines). This median-frequency δ decreases moving up the rows, making discriminability higher on average. Across all panels in a given column, the difference between δ for high-frequency types and δ for low-frequency types (slope) is kept constant. This difference increases (slope steepens) moving rightward across the columns, making high-frequency types increasingly more discriminable than low-frequency types. (B) Results of varying discriminability threshold (δ) with type frequency for representative sets of parameter values (sets (4), (10), and (16) from Table S3 of the Supplementary Materials; all other sets give similar results). The vertical axis shows the extent to which high-frequency types are ahead of low-frequency types in the Pusher (red) or Pushee (blue), averaged over 1000 runs for each parameter setting. A positive slope represents a faster rate of change of high-frequency types compared to low-frequency types. All curves end with a horizontal section corresponding to a stable equilibrium. As in (A), panels are laid out according to δ function. Moving rightward across the columns, high-frequency types become increasingly more discriminable than low-frequency types. This shifts the end of the curve upward for the Pusher (red), causing positive-sloping sections where high-frequency types change at a faster rate than low-frequency types, and the reverse for the Pushee (blue). This effect grows more pronounced moving upward across the rows, as discriminability increases on average. When average discriminability is sufficiently high and high-frequency types are sufficiently more discriminable than low-frequency types (i.e. sufficiently close to the upper-right panel), the model generates robust frequency effects resembling those seen empirically. For discussion of the reverse effects seen when average discriminability is low and high-frequency types are not much more discriminable than low-frequency types (i.e. close to the lower-left panel), see Appendix B. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

frequency effects more generally and compare them to the predictions of other models (Section 6.1). We then highlight the importance of computational modeling in underpinning predictions about sound change and illustrate how further modeling assumptions can generate further testable predictions (Section 6.2). Finally, we discuss the implications of our model for the role of the listener in sound change (Section 6.3).

6.1. General predictions for frequency effects

Unlike previous exemplar-based models of word frequency effects in sound change (Pierrehumbert, 2001, 2002), our model successfully generated different kinds of word frequency effects in different kinds of changes. This success follows from the conception of sound change not merely as the iteration of articulatory biases in the speaker, but rather as the result of balancing emergent forces that stem from both the speaker and the listener, where words of different frequencies are crucially assumed to be differentially sensitive to the perceptual forces in the listener. The force-balancing conception of sound change is entirely general, allowing the model to make predictions for word frequency effects in sound changes beyond the three cases examined here.

The case studies captured by our model can be united under consideration of how each category's change affects its discriminability. When the change has no impact on discriminability – i.e. when the realization of a phoneme drifts phonetically but does not encroach on the acoustic territory of another phoneme – then words of all frequencies are expected to change at the same rate, as in Manchester English /t/-glottaling. When the change acts to decrease discriminability for a category – i.e. when the realization of a phoneme becomes acoustically more similar to that of another phoneme – then high-frequency words are expected to change at a faster rate than low-frequency words, as in New Zealand English /t/-tapping. Finally, when the change acts to increase discriminability for a category – i.e. when the realization of a phoneme becomes acoustically less similar to that of another phoneme – then high-frequency words are expected to change a slower rate than low-frequency words, as in /ɛ/-raising in the New Zealand English short front vowel shift.

By considering the impact on discriminability in other kinds of regular sound change, our model generates a typology of predictions[22] for word frequency effects on rates of change (see the first column of Fig. 10). We believe that the clarity of the predictions, combined with the interesting ways in which they differ for different kinds of sound change, provides a clear and motivated path for future work. Indeed, testing the model's predictions in full will require many empirical studies (see Appendix C), using existing and future corpora of sufficient density and time-depth.[23]

The predictions of our model are particularly interesting because they fit the existing empirical results more closely than those of other approaches. As previously stated, empirical results aligns better with our model's predictions than with the Frequency Actuation Hypothesis (FAH Bybee, 2002; Phillips, 1984),[24] which draws different frequency effects based on the differentiation between articulatory biases and lexical analogy. Empirical results also align better with our model than with a plausible alternative inspired by the FAH (described below), which draws different frequency effects based on the differentiation between articulatory biases and dispersion. In what follows, we work through the logic underlying these three approaches to generate predictions for word-frequency effects in various different kinds of sound change, which we compare schematically in Fig. 10.

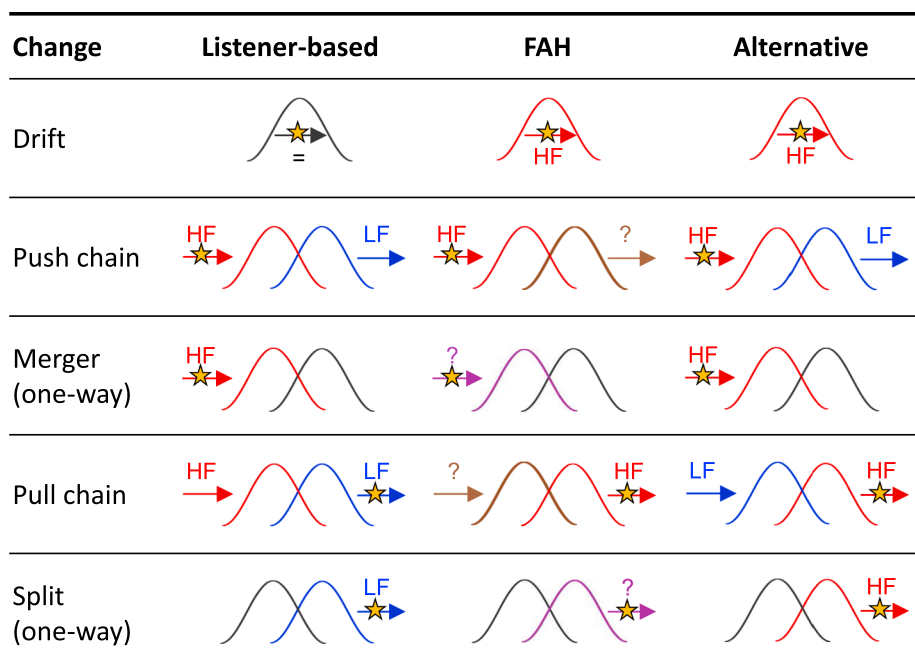As discussed in Section 1, the FAH makes a distinction between

physiologically motivated changes at the surface phonetic level, which are driven by articulatory biases and assumed to affect high-frequency words fastest, and non-physiologically motivated changes at the level of phonological grammar, which are driven by lexical analogy and assumed to affect low-frequency words fastest. While intuitively attractive, this distinction does not make clear predictions for all kinds of sound change. For example, the pushee in a push chain may be argued to move in response to the same articulatory bias that moves the pusher, or in response to dispersive pressures from perception. If the movement is due to articulatory bias, then the FAH predicts a high-frequency advantage, but if it is due to dispersive pressures, then it is unclear what the FAH predicts, as such pressures are based in neither articulation nor analogy. Similarly, sound changes such as merger affect both surface realizations and the phonological grammar and may be driven by articulatory biases or lexical analogy. In such cases, it is unclear whether the FAH predicts a high- or low-frequency advantage.

How can we make clearer predictions than the FAH, while maintaining its core claim that high-frequency words change faster than low-frequency words in physiologically motivated changes and slower in non-physiologically motivated changes? The following alternative approach represents one plausible attempt to do so. Let us assume, as in our established model, that a phoneme category may participate in regular sound change only due to articulatory biases or due to dispersive pressures, and that the number of categories subjected to articulatory biases should be minimized. Thus, for example, the pusher in a push chain moves due to articulatory biases, but the pushee moves due to dispersive pressures. Let us further assume that articulatory bias – a physiological motivator of sound change – affects high-frequency words fastest, while dispersive pressures – non-physiological motivators of sound change – affect low-frequency words fastest. The result is an approach that assumes that sound change occurs in response to pressures from both the speaker and the listener, like our model, but that prioritizes the speaker over the listener, unlike our model. Because any sort of sound change must be motivated *somehow* – whether by biases in production or by pressures from perception – this approach would predict that every sound change should show some word frequency effect. It would predict high-frequency advantages whenever a phoneme is subject to a production bias, and low-frequency advantages in all other movements.

As shown in Fig. 10, the proposed alternative approach makes the same predictions as our model for push chains and mergers, and both are able to account for the observed word frequency effects in /t/-tapping (Hay & Foulkes, 2016) and /ɛ/-raising (Hay et al., 2015) in New Zealand English. We take this observation as support for the general usage-based framework that both approaches take, in which regular sound change occurs in response to forces from both speakers and listeners, due to both articulatory biases and dispersive pressures. However, the predictions differ for phonetic drift, and the empirical observations from /t/-glottaling in Manchester English align with our model's prediction. We take this observation as initial support for our model over the alternative, and correspondingly for treating the listener rather than the speaker as the driver of word frequency effects in regular sound change. Finally, the predictions completely oppose one another for pull chains and splits, but we lack the data at present to test which one is correct. We leave it to future work to test these differing predictions.

6.2. Computational modeling and sound change

The computational model developed in this paper is a significant contribution to the field, both as crucial support for the centrality of the listener to regular sound change, and as an object in its own right. As outlined below, our model has allowed us to remove gaps and uncertainties in predictions by ensuring that they are holistic and internally consistent, to hold intuition up to scrutiny by putting it on a formal foundation, and to clarify the causal relations between

---

[22] Of course, the general predictions of our model are not without limitations; for full discussion, see Appendix C. In addition, model simplifications mean that it cannot be applied to all sound changes; see Appendix A.1.

[23] Within such a corpus, an appropriate analysis could investigate any phonetically gradual change, provided that it aggregates over the entire lexicon. The crucial investigation would be of the statistical interaction between word frequency and time in the prediction of acoustic quality, representing effects of word frequency on rates of change.

[24] It is difficult to make a direct comparison between the predictions of the present model and those of the FAH, since they concern different properties of change (rate versus actuation). To facilitate comparison, we assume neutral initial conditions, i.e. no relevant differences based on word frequency before the onset of the change.

Fig. 10. Comparison of the qualitative predictions of our listener-based model to those of the Frequency Actuation Hypothesis (FAH; Bybee, 2002; Phillips, 1984) and those of our alternative proposal (see text). Arrows indicate movement over time, and stars indicate movement due to phonetic biases. Red indicates a high-frequency advantage, blue indicates a low-frequency advantage, and black indicates no frequency advantage. In the FAH predictions, not every case is clear-cut (marked by ?; see text), introducing cases where there could be a high-frequency advantage or no frequency advantage (brown, no star) or where there could be a high-frequency advantage or a low-frequency advantage (purple, star). The predictions of our listener-based model and the proposed alternative are identical for push chains and mergers, but different for phonetic drift, pull chains, and splits. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

assumptions and predictions. To enable future research to share these benefits, we have made the code for our model available as an online supplement to this paper, together with documentation for how to use it.

Firstly, computational modeling ensures that predictions are holistic and internally consistent. In our model, all sound change is underpinned by movement of exemplars due to forces based in the speaker and the listener, and word frequency effects are driven by listener-based forces, in entirely general ways. Since every kind of sound change can be conceived of in relation to these forces, the model makes clear (probabilistic) predictions for every kind of sound change (modulo the limitations discussed in Appendix C). The same is not true of previous hypotheses that have not been implemented in computational models, such as the FAH (see Section 6.1). Because the FAH only addresses sound changes caused by articulatory biases or lexical analogy, it does not extend to changes which do not seem to be caused by either (e.g. the pushee in a push chain), and it makes unclear predictions for changes which could plausibly be caused by both (e.g. merger).

Because our model is a formal implementation, it allows us to hold intuition up to to scrutiny. For example, in Section 4.2, we showed that our model does not support the widespread intuition that exemplar-based models always predict high-frequency words to change fastest in response to production biases (Abramowicz, 2007; Bermúdez-Otero et al., 2015; Dinkin, 2008; Tamminga, 2014). This intuition relies in part on a conflation of type frequency and category frequency. We represent types and categories at separate levels, making it clear that type frequency affects not only rate of production, but also density of exemplar distribution. These two effects counteract one another in determining how quickly the type moves in response to bias (see also Section S5.2 of the Supplementary Materials for discussion from a mathematical point of view).

Finally, our model has allowed us to clarify the causal relations between assumptions and predictions. By showing that the model predicts different frequency effects with the assumption of a high-frequency discriminability advantage (Section 5.5) than without it (Section 5.2), we have established that word frequency effects on rate of change can be causally related to word frequency effects on perception. The model also enables us to establish the influence that changing certain assumptions would have on predictions. For example, there are at least two possible mechanisms through which the existence of

asymmetries in discriminability evaluation can follow from the exemplar-based architecture of the model (i.e. under a constant discriminability threshold, δ); in Section S1.3.1 of the Supplementary Materials, we outline these different mechanisms and show how they make different predictions for asymmetries in typicality evaluation.

6.3. Listener-driven sound change

Our model is formulated as a usage-based production-perception loop, where the representations that are drawn upon for production are also updated through perception. Within such a system, the diachronic trajectory of a language is formed from the way in which the language is used at different points in time, and thus shaped by the forces (social, cognitive, physiological, etc.) that act during any synchronic linguistic exchange (Beckner et al., 2009). More generally, the interdependence between the production and perception systems predicts that *any* sort of synchronic asymmetry in the way speech is produced or perceived has the potential to create asymmetries in the pattern of diachronic change, provided it is sufficiently widespread, robust, and persistent. By consequence, it opens the possibility that the listener could drive regular sound change.

The notion that the listener could be important for sound change is not new, but our approach to it is. Ohala (1981) claimed that the listener could be a source of sound change by under- or over-applying perceptual compensation for coarticulation and thus misperceiving one sound as another. For example, the listener could incorrectly compensate for coarticulation that wasn't present in the realization of /yt/ as [yt] and thus reconstruct it as /ut/. There are three main differences between Ohala's model and the model we present here. Firstly, Ohala's model represents sound change as a change in the phonological representation of words (e.g. /yt/ changing to /ut/), while our model represents sound change as a change in the phonetic realization of a phoneme (e.g. /y/ changing its realization from [y] to [u]). Consequently, Ohala's model is not designed to capture the gradient changes that are observed in regular sound change, while our model is. Secondly, Ohala's model attributes the influence of the listener to misperceptions, while our model attributes it to memory disadvantages of acoustically ambiguous tokens. While listeners certainly can misperceive one sound as another, especially in the presence of noise (Miller & Nicely, 1955), we do not believe that misperception is as

widespread in practice (especially given context) as would be required for it to really drive sound change (see Section S3.4.3 of the Supplementary Materials for related simulations showing that misperception of minimal pairs is insufficient to generate robust category interaction). Finally, in Ohala's model, a sound change that spreads across the lexicon (outside of the conditioning environment) must do so via analogy – meaning that the listener is a *source* of sound change, but does not *drive* sound change across the lexicon. In our model, the speaker is the source of sound change (via biases in production), but the listener drives it, by forcing categories to interact (via the discriminability evaluation).

Some models of language usage acknowledge the interdependence of production and perception while still placing primary focus on the speaker. Such models typically involve a forecasting component, according to which speakers shape their productions to facilitate comprehension of a listener (e.g. Buz et al., 2016; Lindblom, 1990).[25] While consistent with all we have demonstrated for the listener-based model, these speaker-based, listener-oriented models have two primary disadvantages. Firstly, it is not clear that a listener-oriented component is consistently active; for example, while Snedeker and Trueswell (2003) observe speakers' use of prosodic cues covarying with – and thus potentially disambiguating – prepositional phrase attachment, Arnold et al. (2004) do not. Similarly, Bard et al. (2000) observe that speakers reduce intelligibility when repeating referring expressions, even when they know that the listener has no knowledge of the referent. Secondly, listener-orientation has been criticized as teleological (e.g. Wedel, 2006). By contrast, the present listener-based model is non-teleological: change occurs not because of any desire to speak in an acoustically unambiguous manner, but because instances of ambiguous speech are disfavored in memory and thus tend not to be replicated, following entirely passive processes.

Our results have shown that a listener-based approach to sound change explains empirical patterns that cannot be satisfactorily captured by existing speaker-based approaches. By locating synchronic word frequency-based asymmetries in the perceptual system, we have predicted frequency effects on rate of change that differ depending on the kind of sound change in action. Our predictions capture empirical results from /t/-glottaling in Manchester English (Bermúdez-Otero et al., 2015) and /t/-tapping (Hay & Foulkes, 2016) and /ɛ/-raising (Hay et al., 2015) in New Zealand English, which, to our knowledge, have no other joint explanation in the literature. Of course, there are many perceptual asymmetries that we have not included in our model, and our approach leads us to expect that they may also have implications for patterns of sound change.

In sum, the listener-based approach is both powerful and flexible. We take the success of the approach in predicting empirically-observed effects of word frequency on rate of change to support the claim that the listener is central to sound change. We do not intend this to claim imply that the speaker is unimportant, but rather that the speaker cannot be the sole primary influence on sound change. We acknowledge that aspects of production are widely attested and accepted to vary with word frequency, at least in the case of reduction (e.g. Bell et al., 2009). However, given an exemplar-based production-perception loop, it is not necessary to assume that word frequency-based asymmetries in production are responsible for generating effects of word frequency on rate of change, as such effects can follow straightforwardly from frequency-based asymmetries in perception.

7. Conclusion

We have presented an exemplar-based computational model of regular sound change and demonstrated that it generates appropriate single-category movement and two-category interactions, reflecting key (frequency-independent) properties of real sound changes. In particular, the model is capable of maintaining substantial overlap between phoneme category distributions, owing in part to the novel inclusion of an experimentally-supported typicality evaluation in perception.

We have then shown that the model's assumption that the listener plays a central role in sound change allows it to predict different effects on word frequency on rate of change in different kinds of sound change, which match all of the empirical results that exist at the time of writing (Bermúdez-Otero et al., 2015; Hay et al., 2015; Hay & Foulkes, 2016). In changes that do not affect the acoustic ambiguity of the phoneme undergoing change, the model predicts all words to change at the same rate, owing to the novel de-confounding of word frequency and phoneme frequency. In changes that act to increase the acoustic ambiguity of the phoneme undergoing change, the model predicts high-frequency words to change at a faster rate than low-frequency words, and vice versa for changes that act to decrease the acoustic ambiguity of the phoneme undergoing change. These predictions follow from the incorporation of an experimentally-supported perceptual asymmetry, under which high-frequency words may be more robustly recognized than low-frequency words in the face of acoustic ambiguity. Our listener-based model thus shows how word frequency-based asymmetries in perception can generate word frequency effects on rate of sound change, without similar asymmetries in production, which differ systematically for different kinds of sound change.

The study of regular sound change is becoming increasingly rigorous with the availability of collections of speech recordings that span long periods of (real or apparent) time. The combination of this empirical data with appropriate computational modeling will be central to testing predictions and hypotheses about the connections between speech perception and regular sound change. Under the model we have presented here, we expect future studies of empirical data to demonstrate connections between word frequency and rate of sound change, and we expect these connections to differ systematically depending on the implications of the change for perceptual discriminability.

[25] A reviewer points out an alternative "speech-monitoring" model, in which the speaker does not actively design their speech for a listener, but rather adjusts their speech based on their own perceptual processes prior to production. While the details of such a model remain to be worked out, it is a plausible alternative. Crucially, it is not inconsistent with our main point, as it still implies a central role for perceptual processes. However, it can only be indirectly supported by results from perception experiments, whereas our model can be directly supported.

Appendix A. Simplifications in modeling decisions

In order to develop a thorough understanding of the fundamental dynamics of our model, we made a number of necessary simplifications. In this section, we describe those simplifications and their implications for model behavior and applicability to empirical sound change data.

A.1. The exemplar space

Our model uses an exemplar space that is subject to three assumptions: (i) it contains just one or two categories; (ii) it consists of a single perceptual-acoustic dimension; and (iii) it extends without bound, with all areas being equally "hospitable" for exemplars. Relaxing any of these assumptions would introduce complexities that are beyond the scope of the present work.

Relaxing the first assumption – allowing more than two categories into the model – would cause the predicted frequency effects to depend upon the precise initial configuration of categories. For example, in a system that includes three categories along a single dimension, different effects will be predicted based on whether the middle category begins closer to the first or last category.

Relaxing the second assumption – expanding to a multi-dimensional exemplar space – would cause the predicted frequency effects to depend upon the alignment of the category trajectories. For example, if the Pusher is moving horizontally, different frequency effects will be predicted based on whether the Pushee also moves horizontally or instead moves with a vertical component.

Finally, relaxing the third assumption – adding bounds to the exemplar space that repel nearby exemplars, e.g. via physical limits of articulation – would cause the predicted frequency effects to depend upon the role of those bounds in the change. For example, in a two-dimensional system,[26] different frequency effects will be predicted based on whether Pushee movement along a certain dimension is caused by repulsion from the Pusher or from the bounds of the space.

As a result of these assumptions, our present model is only applicable to sound changes that can be approximated by the movement of one or two categories along a single trajectory. For some complex sound changes, it is possible to identify simple components that can be approximated in this way. For example, the New Zealand short front vowel shift as a whole (Hay et al., 2015) cannot be captured by our present model, as it involves at least three categories moving around the boundaries of a two-dimensional space. But within this complex change, one component – /ɛ/-raising – is amenable to treatment by our model, as it involves movement of a category (/ɛ/) along the same trajectory as its pusher (/æ/) but not its pushee (/ɪ/). We plan extension of the model to other components of the change in future work.

A.2. Minimal pairs

The simulations presented in the main paper do not include minimal pairs, under the claim that minimal pairs alone are unlikely to drive the effects that we are concerned with modeling. We support this claim in two ways: by showing that minimal pairs are in a minority in empirical sound changes, and by showing via simulations that a minority of minimal pairs is neither sufficient nor necessary to generate behavior of interest in a model of sound change.

In the interaction of /æ/ and /ɛ/ in the New Zealand short front vowel shift, only 8.2% of words (164 of 2000 unique wordforms) have a relevant non-proper-noun minimal partner that also appears in the ONZE corpus. These minimal pairs are distributed across the frequency range and account for 21.1% of the total tokens (11,620 of 55,200) analyzed by Hay et al. (2015). Since the vast majority of the New Zealand English /æ/-/ɛ/ data (words and tokens) correspond to words without a relevant minimal partner,[27] we argue that the properties of the vowel interaction are likely to be general, holding across words both with and without relevant minimal partners.

We further support this argument with simulations, detailed in the Section S3.4 of the Supplementary Materials. If we assume that a minority of minimal pairs alone are responsible for generating phoneme category interaction, our model produces little interaction and no word-frequency effects. Conversely, if we assume that phoneme category interaction is general, i.e. not generated solely by minimal pairs, our model produces substantial interaction and word-frequency effects. Under this assumption, the model produces the same qualitative results independent of whether minimal pairs are included or excluded. By excluding minimal pairs, we thus do not exclude any behavior of interest from our model, and we guarantee that the predictions of our model are general.

A.3. Frequency

Our model uses a representation of type frequency that is based on log-transformed word frequencies in a large corpus (see Section S1.1 of the Supplementary Materials for motivation and discussion). In addition to using this representation for modeling perceptual processes (as is standard), we use it for modeling rates of production. This treatment might seem inappropriate because, in the real world, words are produced according to their frequency rather than their log-frequency. However, it is not, because our model is fundamentally concerned only with productions *that are*

---

[26] In a one-dimensional system, a bounded exemplar space would cause movement to cease eventually, but would not otherwise interact with frequency effects. This lack of interaction allows our model to be applied to movement that eventually ceases – as in /t/-glottalling and -tapping – even though movement never actually ceases in simulations.

[27] The number of words in the dataset with potential relevant non-proper-noun minimal partners (as assessed by the Unisyn lexicon (Fitt, 2000)) that happened not to be mentioned (e.g. through not being of the appropriate register or through being extremely low-frequency) is capped at 15.4% (308), accounting for no more than 37.1% (20,501) of the tokens. Even in this more extreme interpretation of the data, minimal pairs are in a minority.

assessed for storage in memory, which need not include all words in a stream of speech (see also Landauer, 1986, for a model of memory in which not every input is stored). Similarly to recent subsampling approaches in Natural Language Processing (Mikolov et al., 2013), we assume that listeners may filter out (or otherwise downweight) some instances of high-frequency words due to their high predictability. We take the liberty not to model such filtered instances for the sake of computational efficiency.

A.4. Exemplar storage and strength

Our model assumes that a stored exemplar overwrites another, and thus that all exemplars have a fixed strength that does not decay over time. This treatment is different to the standard one in exemplar dynamics models, in which there is no overwriting, but exemplar strength decays exponentially over time (Ettlinger, 2007; Pierrehumbert, 2001; Tupper, 2015; Wedel, 2006, 2012; Wedel & Fatkullin, 2017). However, the difference is superficial; averaged over many runs, the expected behavior of our random-overwriting treatment is equivalent to that of a special case of the standard treatment (see Sections S5.2.1–5.2.2 of the Supplementary Materials).

Our treatment affords us a computational convenience, as the number of exemplars in each category remains fixed over time. It also has the consequence that one category cannot "leech" off the other, driving it to extinction. While previous work (e.g. Pierrehumbert, 2001; Tupper, 2015) has used category extinction via leeching as a model for phoneme merger, Wedel and Fatkullin (2017) show that such leeching behavior is naturally avoided by even partial grounding of production rates in the lexicon, meaning that it may not be entirely appropriate as a model of merger. In the absence of leeching, we treat merger as the result of extreme overlap between categories (Sóskuthy, 2013; Wedel, 2004, 2012), following the assumption that such overlap may trigger reanalysis of the category system cross-generationally (Blevins, 2006) or updating of lexical entries within a generation.

Appendix B. Interactions between production and storage

Our simulations contained as many exemplars of high-frequency types as exemplars of low-frequency types (see Section S1.1 of the Supplementary Materials). In the absence of perceptual asymmetries, the high- and low-frequency sub-distributions of exemplars are therefore expected to evolve identically. However, that is not what happened, for both the single-category simulations and the two-category simulations. In this section, we describe the unexpected frequency effects, and show how they are explained by a side-effect of our assumptions about production and storage.

Our single-category simulations revealed an effect of type frequency on the evolution of category width. Though the category as a whole maintained its width throughout the simulations, the sub-distribution corresponding to high-frequency types narrowed and the sub-distribution corresponding to low-frequency types widened. We demonstrate this result in Fig. B.11.
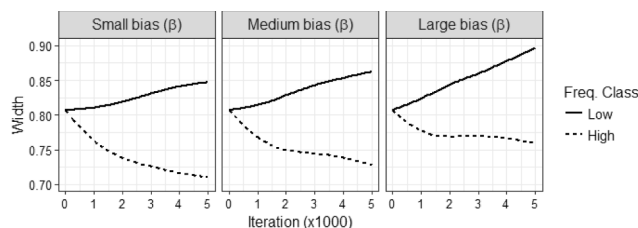
Fig. B.11. Results of simulations with a single category ($\sigma = 0.8$) subject to varying degrees of bias, illustrating differences between low-frequency (solid) and high-frequency (dashed) types. For all degrees of bias, the category distribution widens for low-frequency types, but narrows for high-frequency types. Category skewness increased with bias (β), causing apparent category width also to increase.

Similarly, our two-category simulations with the basic model (i.e. without frequency-sensitive discriminability threshold) revealed an effect of frequency on rate of change that was the opposite of that seen empirically (Fig. 8). Low-frequency types changed fastest in the Pushee, and high-frequency types changed fastest in the Pusher.

Both of these results are conditioned by perceptual forces: category width is conditioned by the typicality force, and position in two-category interactions is conditioned by the discriminability force. In both cases, the high-frequency sub-distribution has been more affected by the relevant force than the low-frequency sub-distribution. In other words, high-frequency types are more *sensitive* to perceptual forces that low-frequency types.

High-frequency types have increased sensitivity to perceptual forces in general as an indirect result of an interaction between our assumptions about production and storage. Recall that the selection of an initial target for a token proceeds by copying the acoustic value of an exemplar of the given type (Section 3.2.1), and that a token is unlikely to be stored if it falls in a perceptually-disadvantaged (i.e. low-discriminability or low-typicality) area of the exemplar space (Sections 3.2.6 and 3.2.7). Consequently, an exemplar in a perceptually-disadvantaged area (e.g. an atypical exemplar) is less likely to generate a token that will overwrite an exemplar in a perceptually-advantaged area (e.g. a typical exemplar) than vice versa. In this way, the target-copying production mechanism provides escape routes from perceptually-disadvantaged areas of the exemplar space, yielding boosts to perceptual forces. An exemplar of a high-frequency type falling in a perceptually disadvantaged area has more escape routes than a similar exemplar of a low-frequency type because its representation consists of more exemplars, and thus gets more boosts to the perceptual forces.

We emphasize that this sensitivity difference is a side-effect of model assumptions and should not be interpreted as theoretically meaningful. None of our key results rely on it in any way.

Appendix C. Limitations on testing model predictions

Our model predicts a typology of word frequency effects on rates of sound change (Section 6.1), which can be tested with corpus data. However, these predictions are not without limitations. In addition to the limitations already discussed in Appendix A.1 on the kinds of changes to which the

(A)

(B)
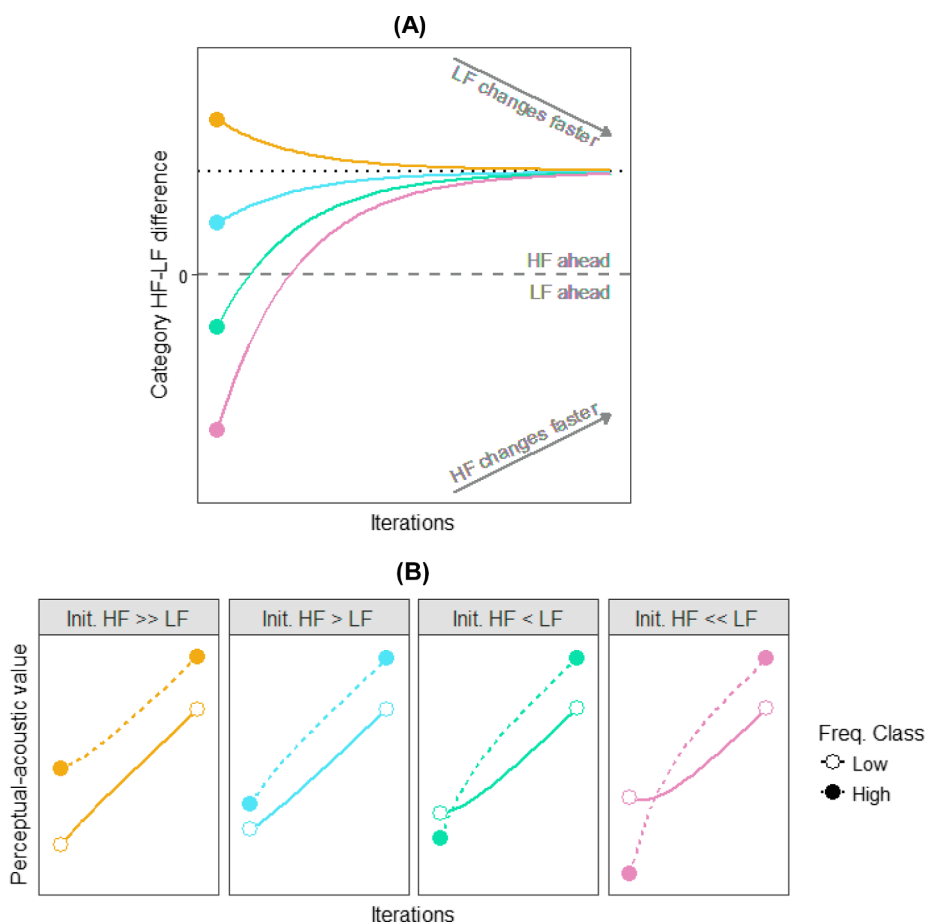
Fig. C.12. Illustration of how the model predictions work, assuming that the sound change is causing increase along some perceptual-acoustic dimension. (A) illustrates change in the difference along the perceptual-acoustic dimension between high- and low-frequency subdistributions of the category over time, while (B) illustrates potential trajectories of sub-distribution movement that could generate the patterns in (A). Both illustrations include possibilities from different initial conditions (matched colors), i.e. different relative positions of the high- and low-frequency subdistributions at the beginning of the change. The model predicts eventual convergence to a fixed difference (dotted black line in (A); fixed separation of trajectories in (B)). This prediction is not robustly related to a prediction of which types are "ahead" in the change at any given point in time. Half of the initial conditions (orange, blue) have high-frequency types ahead (above the dashed gray line in (A); higher value of the high-frequency trajectory in (B)) at all points in time, while the other half (green, purple) have low-frequency types ahead at first and high-frequency types ahead later. The model's prediction is more robustly related to a prediction of differences in rates of change. Most of the initial conditions (blue, green, purple) have high-frequency types change faster than low-frequency types (positive slope in (A); steeper slope of the high-frequency trajectory in (B)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

predictions are applicable, there are two main limitations on the ways in which the predictions can be tested. In this section, we describe these limitations.

The first limitation is that the force-balancing conception ultimately concerns the *eventual* behavior of a system, which cannot be directly investigated in a corpus. The model predicts the relative positions of high- and low-frequency types upon converging to a stable equilibrium state where all forces are balanced[28]; we illustrate how this prediction works in Fig. C.12. In a corpus, one cannot directly test for convergence to such an equilibrium state because of the inherent difficulty of saying when a change has reached stability, due to the existence of substantial variation. Consequently, the model's predictions can only be compared to corpus results indirectly. We have focused our exposition of the model on the effect of type frequency on rates of change, which can be observed in corpora as a statistical interaction between word frequency and phoneme realization (Hay et al., 2015). The effect of frequency on rates of change is a fairly robust indirect prediction of the model that holds across most initial conditions (see Fig. C.12); however, it may not be appropriate in some (extreme) circumstances, depending on the initial conditions.

The second limitation is that, because the model is stochastic in nature, its predictions reflect what is expected *on average* for a particular kind of change, not what actually happens in any *instance* of that change. The model's predictions are tendencies that should be observable *across* instances of change, and thus can only truly be tested by analyzing many different corpora. Even under the model settings that generated the strongest frequency effects in the exploration presented here (top-right panel of Fig. 9), only 70% of our simulations yielded the corresponding qualitative pattern of results, some of which represent extremely small effect sizes. Thus, the existence of a certain number of null or even conflicting results from empirical studies is not inconsistent with the model, but we expect that, over many studies, more results will show the predicted qualitative patterns than will show any other pattern. Similarly, the model's predictions for frequency-based differences across types within a particular change are also tendencies. The model does not predict that *every* pair of high- and low-frequency types within a category will exhibit the relationship that is expected for the given change; rather, the relationship is only expected to manifest when aggregating over the entire lexicon (or a sufficiently large representative sample).

Appendix D. Supplementary material

Supplementary materials and model code associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cognition.2019.01.004.

---

[28] In the eventual force-balance equilibrium state, the categories are still moving, but the movement of types of different frequencies neither accelerates nor decelerates, like a car at stable cruising speed. Because the model assumes that the representations (e.g. lexical and phonological inventories) and forces (e.g. production bias) are constant for all time, it is always driven toward a single equilibrium state, even if this state is not reached within the specified iterations. In real sound systems, the representations and forces may change readily, meaning that "convergence" never actually occurs, and instead the expected frequency effects change with the system.

References

Abramowicz, Ł. (2007). Sociolinguistics meets exemplar theory: Frequency and recency effects in (ing). *University of Pennsylvania Working Papers in Linguistics, 13*, 27–37.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language, 51*, 55–70.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning, 59*, 1–26.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92–111.

Bermúdez-Otero, R., Baranowski, M., Bailey, G., & Turton, D. (2015). A constant rate effect in Manchester /t/-glottalling: High-frequency words are ahead of, but change at the same rate as, low-frequency words. *Paper presented at 2nd Edinburgh symposium on historical phonology*.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics, 32*, 117–166.

Blevins, J., & Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica, 26*, 143–183.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707–729.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language, 89*, 68–86.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*, 261–290.

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics, 70*, 604–618.

Clopper, C. G., Pierrehumbert, J. B., & Tamati, T. N. (2010). Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology, 10*, 65–92.

Clopper, C. G., Tamati, T. N., & Pierrehumbert, J. B. (2016). Variation in the strength of lexical encoding across dialects. *Journal of Phonetics, 58*, 87–103.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*, 234–250.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 81–94.

de Courtenay, B. (1895). *Versuch einer Theorie phonetischer Alternationen: ein Capitel aus der Psychophonetik*. Strassburg: Trübner [English translation by E. Stankiewicz (1972). An attempt at a theory of phonetic alternations. In E. Stankiewicz (Ed.), A Baudoin de Courtenay anthology: The beginnings of structural linguistics (pp. 144 –212). Bloomington, IN: Indiana University Press.].

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition, 108*, 710–718.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*, 317–367.

Dinkin, A. J. (2008). The real effect of word frequency on phonetic variation. *University of Pennsylvania Working Papers in Linguistics, 14*, 97–106.

Emanuel, R., Adams, J., Baker, K., Daufin, E. K., Ellington, C., Fitts, E., ... Okeowo, D. (2008). How college students spend their time communicating. *International Journal of Listening, 22*, 13–28.

Ettlinger, M. (2007). An exemplar-based model of chain shifts. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 685–688).

Fitt, S. (2000). *Documentation and user guide to unisyn lexicon and post-lexical rules. Technical report*. Centre for Speech Technology Research, University of Edinburgh.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.). *The phonetic bases of markedness*. Cambridge, UK: Cambridge University Press.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627–635.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*, 396–413.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 526–540.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*, 474–496.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110–125.

Garrett, A. (2015). Sound change. In C. Bowern, & B. Evans (Eds.). *The Routledge handbook of historical linguistics* (pp. 227–248). New York, NY: Routledge.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access.

Psychological Review, 105, 251–279.

Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2018). Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science*, 1–22.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 4*, 865–892.

Hay, J., & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language, 92*, 298–330.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*, 458–484.

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition, 139*, 83–91.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*, 3099–3111.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology, 88*, 297–306.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *The Journal of the Acoustical Society of America, 29*, 296–305.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. Mullenix (Eds.). *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*, 485–499.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*, 148–203.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*, 262–268.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Labov, W. (2010). *Principles of linguistic change, Vol. 3: Cognitive and cultural factors*. Oxford: Wiley-Blackwell.

Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science, 10*, 477–493.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*, 839–862.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.). *Speech production and speech modelling* (pp. 403–439). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear, 19*, 1–36.

de Marneffe, M.-C., Tomlinson, J., Tice, M., & Sumner, M. (2011). The interaction of lexical frequency and phonetic variation in the perception of accented speech. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.). *Proceedings of the 33rd annual meeting of the Cognitive Science Society* (pp. 3575–3580). Austin, TX: Cognitive Science Society.

Martinet, A. (1952). Function, structure, and sound change. *Word, 8*, 1–32.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., Chen, K., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.). *Proceedings of Advances in Neural Information Processing Systems 26* (pp. 3111–3119). Curran.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338–352.

Milroy, J., Milroy, L., Hartley, S., & Walshaw, D. (1994). Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English. *Language Variation and Change, 6*, 327–357.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*, 62–85.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*, 132–142.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–61.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.). *Papers from the parasession on language and behavior* (pp. 178–203). Chicago, IL: Chicago Linguistic Society.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119*, 2382–2393.

Phillips, B. S. (1981). Lexical diffusion and Southern tune, duke, news. *American Speech, 56*, 72–78.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language, 60*, 320–342.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds.). *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.). *Laboratory Phonology VII* (pp. 101–139). Berlin: Mouton de Gruyter.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 699–725.

Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *The Journal of the Acoustical Society of America, 35*, 200–206.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103–130.

Sóskuthy, M. (2013). *Phonetic biases and systemic effects in the actuation of sound change* (Unpublished doctoral dissertation)University of Edinburgh.

Sóskuthy, M. (2014). Explaining lexical frequency effects: A critique and an alternative account. *Paper presented at sound change in interacting human systems, 3rd biennial workshop on sound change.*

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.). *Natural language processing and speech technology: Results of the 3rd KONVENS conference* (pp. 14–26). Berlin: Mouton.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology, 4*.

Tamminga, M. (2014). Sound change without frequency effects: Ramifications for phonological theory. In R. E. Santana-LaBarge (Ed.). *Proceedings of the 31st West Coast Conference on Formal Linguistics* (pp. 457–465). Somerville, MA: Cascadilla Proceedings Project.

Tupper, P. F. (2015). Exemplar dynamics and sound merger in language. *SIAM Journal on Applied Mathematics, 75*, 1469–1492.

VanDam, M. (2007). *Plasticity of phonological categories* (Unpublished doctoral dissertation)Indiana University.

Wedel, A. (2004). Category competition drives contrast maintenance within an exemplar-based production/perception loop. In J. Goldsmith, & R. Wicentowski (Eds.). *Proceedings of the seventh meeting of the ACL special interest group in computational phonology* (pp. 1–10). Stroudsburg, PA: Association for Computational Linguistics.

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

Wedel, A. (2012). Lexical contrast maintenance and the development of sublexical contrast systems. *Language and Cognition, 4*, 319–355.

Wedel, A., & Fatkullin, I. (2017). Category competition as a driver of category contrast. *Journal of Language Evolution,* 77–93.