VESTIGE THEORY: SOCIOLINGUISTIC EVIDENCE FOR OUTPUT-OUTPUT CONSTRAINTS[1]

Abstract

This article proposes a model of variation and change within the Optimality Theory framework. Previous attempts to account for variation in Optimality Theory have a common flaw; they fail to include a link to the external world where demographic factors such as gender, age, style, register, and social class play a material role in linguistic behavior. I propose Vestige Theory as a means of using Optimality Theoretic architecture to license the influence of language external factors. Vestige Theory claims that language change is always the result of constraint demotion and that the demotion of a constraint leaves behind a vestige of itself. This vestige constraint is a kind of output-output constraint. It differs from the "true" constraint that is demoted as part of language change. As an output-output constraint, a vestige constraint has the link to the outside world (i.e. to external factors) that previous models have failed to incorporate.

Keywords: Variation, Sound Change, Optimality Theory

1. Introduction

Traditional Optimality Theory (Prince and Smolensky 1993), henceforth OT, has often been an effective tool for capturing the internal factors of language (phonology, morphology, etc.). However, the field of sociolinguistics has demonstrated that external factors (gender, age, style, register, and social class) are almost always involved in linguistic behavior as well. Because traditional OT is not adept at integrating external factors into the model, describing variation in OT remains a challenge. The problem is described by Anttila (2002: 212):

> "How do internal factors interface with external factors in variation? While it is not the business of grammatical theory to explain the effects of sex, age, style, register and social class, one would like to have at least a plausible scenario of where such facts fit. There would seem to be two possibilities: (1) The modular view: internal and external factors are of a fundamentally different nature: grammars are structural objects built out of (innate) universal principles; external factors reflect the ways in which these structural objects are used. This implies that external factors can be reduced to choices among grammars; (2) The anti-modular view: there is no important theoretical difference between internal and external factors, which interact with each other fairly directly."

In this article I propose an anti-modular view of the external factors that condition variation. I will refer to this proposal as Vestige Theory. Vestige Theory claims that language change is always the result of constraint demotion and that the demotion of a constraint leaves behind a vestige in its original position. A vestige constraint is therefore the phantom of the demoted constraint. Crucially, the vestige constraint is a

[1] I am grateful to Tracy Alan Hall for his invaluable feedback throughout the development of the paper. Thanks also to Julie Auger, Kari Ellen Gade and Rex Sprouse for their helpful comments to improve earlier drafts. Additionally, I thank the audience at MCWOP 14 for valuable discussion. All errors are my own.

kind of output-output constraint and therefore differs from the "true" constraint that was demoted. As an output-output constraint, a vestige constraint has the link to the outside world (i.e. to external factors) that all previous methods of describing variation in OT have failed to incorporate.

The kind of output-output model that is employed in this article is derived from the Base-Identity model used in explaining phenomena of Paradigm Uniformity (cf. Downing *et al.* 2005 and sources therein). Specifically, I argue that an individual is a cell of a social paradigm and that the norms of that social paradigm cause the individual to produce contextually appropriate outputs that are not surface true. This is analogous to the influence that a base has on a grammatical paradigm.

The article is organized as follows: in section 2, I discuss the problem of variation in OT. I argue that no treatment of variation in OT will ever make correct predictions about the external factors that condition variation. Rather, a phonological model of variation needs to *license* the role of such external factors. In section 3, I turn to an example of a paradigm uniformity effect. The purpose of this example is to demonstrate how output from the external world sometimes interferes with the selection of a particular candidate. Specifically, I illustrate how this interference is predicted using a Base-Identity model. In section 4, I derive Vestige Theory from the Base-Identity model described in section 3. Section 5 examines the main assumptions of the proposed model through an analysis of Norwich develarization (sometimes called "g-dropping"). Section 6 examines leading models of variation and concludes that these models will always be insufficient because they never make reference to the output of the external world. Finally, section 7 is a prospectus in which I suggest possible directions for future research.

2. The Problem of Variation in Optimality Theory

Variation has always posed a problem for OT. Some of the earliest accounts of variation within the OT framework were concerned with defining dialect grammars. In this sort of modeling of variation, each grammar remains separate and distinct from the other. McCarthy (1993: 169), for example, examines the following data from Bostonian English to argue for three dialects:

(1)

Dialect A	Dialect B	Dialect C
Wanda left	Wanda left	Wanda left
Homer left	Home<r> left	Home<r> left
Wanda arrived	Wanda arrived	Wanda[r] arrived
Homer arrived	Homer arrived	Homer arrived

These data indicate that two processes, r-deletion and r-insertion, distinguish the three dialects. In Dialect A, there is neither deletion nor insertion. Dialect B is non-rhotic (<r> indicates an r that is not pronounced), however it does not have any r-epenthesis. Finally, Dialect C shows both processes ([r] indicates the inserted rhotic).

Anttila and Cho (1998) adopt the same[2] constraints proposed by McCarthy (1993)[3]:

[2] The constraints have slightly different names, but for all intents and purposes are the same. Thus, McCarthy's FILL constraints correspond to Anttila and Cho's FAITH constraint, etc.

(2) FAITH 'No deletion, no insertion'
*CODA 'Syllables do not have codas'
ONSET 'Syllables have onsets'

FAITH militates against both Bostonian rhotic phenomena, r-deletion and r-insertion. *CODA provides a motivation for r-deletion and ONSET provides an argument for r-insertion. Accordingly, the constraint rankings in (3) predict the data for the three dialects in (1).

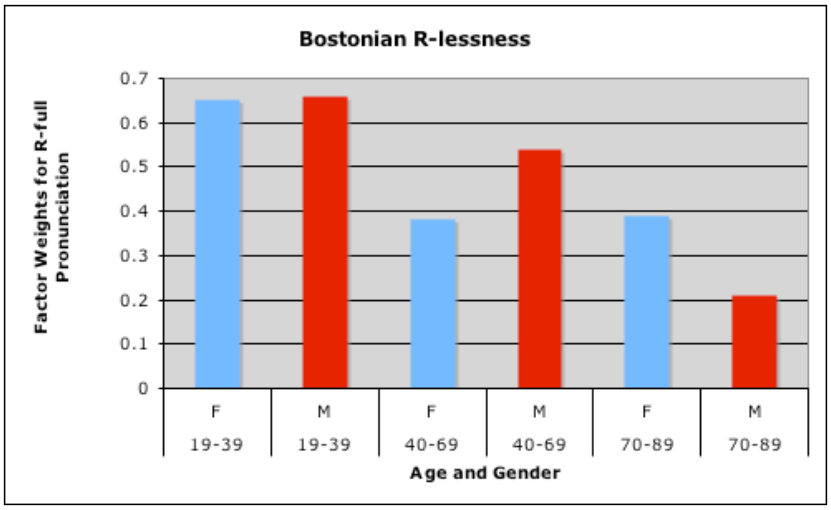
(3)

	Constraint Ranking
Dialect A:	FAITH >> *CODA >> ONSET
Dialect B:	*CODA >> FAITH >> ONSET
Dialect C:	*CODA >> ONSET >> FAITH

McCarthy's original argument is criticized by sociolinguists (cf. Hay and Sudbury 2005; Irwin and Nagy 2007) because it implies that the dialectal pronunciations are categorical for an individual. That is, speakers of Dialect A *never* delete and *never* insert r-sounds, just as speakers of Dialect C, given the appropriate context, *always* delete and *always* insert r-sounds. Critics point out that the true nature of the Bostonian dialect is a mix of all three grammars.

Irwin and Nagy (2007: 144) find the following distribution of r-full and r-less pronunciation in Bostonian English:

(4)



In (4) there are two sociolinguistic factors that play a role in the variability between r-full and r-less pronunciation[4]. One factor is age; the other is gender. The oldest male speakers exhibit the

[3] There is room for further research on the subject since these accounts are somewhat problematic in capturing the fact that the /r/ in a complex coda is more likely to vocalize (Irwin and Nagy 2007); Note also that Labov's (1966, 1972) accounts of New York City r-lessness reports higher rhoticity in *floor* than in *fourth*. Future research needs to consider these data more closely.

[4] Note that the "Factor Weights" in (7) are obtained through multivariate analysis. Tagliamonte (2006: 145) explains, "When a factor weight is closer to 1, it is interpreted as 'favoring' the application value [r-full pronunciation], whereas if it is closer to 0 it is interpreted as 'disfavoring' the application value [r-less pronunciation]."

most non-rhotic speech; the youngest speakers, male and female, are about equally inclined to pronounce /r/ in words like *fourth*. What is important to draw from (4) is that *extra-linguistic* factors play a role in the extent to which one form is chosen over another.

The involvement of extra-linguistic factors adds a lot of complexity. Consider, for example, the data reported in Bell (1991). Speakers were found to use different variants depending on the listening audience. Specifically, New Zealand radio broadcasters sometimes voice the intervocalic /t/ in words like *writer* and *latter*. When broadcasting to a national audience with higher social status, newscasters more frequently used the prestige form with [t]. However, when the same speakers broadcast to local channels, they systematically use the less prestigious [d] pronunciation with greater frequency.

Since an OT tableau represents the cognition of a single person, the use of tableaux without some kind of reference to output in the environment surrounding the individual will *always* fail to capture the principled heterogeneity (Weinreich *et al.* 1968) that is characteristic of all the world's languages. For this reason, I turn to a Base-Identity model of OT where output is assumed to play a role in candidate selection.

In section 3, below, I give an example of a traditional Base-Identity analysis. Then in section 4, I argue that this model is productive for understanding variation in OT because it provides the crucial link to the outside world– to output– that is missing from the other models of variation that are discussed in section 6.

3. An Example of Base-Identity and the Effectiveness of Output-Output Constraints

In the previous section, I claimed that an output-output model of OT would be adept at describing variation because it references the environment surrounding the individual speaker. The purpose of this section is to give an example of Paradigm Uniformity (cf. Downing *et al.* 2005 and sources therein). At first glance, a paradigm uniformity effect has ostensibly nothing to do with the variation problem described in section 2. This is not the case, however. I argue in section 4 that an individual is part of a social paradigm and, as such, is pressured to conform to the output norms of the social group to which that individual identifies: this pressure to conform is analogous to the influence that the base has on the behavior of a cell of a grammatical paradigm. For this reason I turn to some data from Ingveonic[5].

In Ingveonic languages, nasals in vowel-nasal-fricative (VnF) sequences were deleted. This was accompanied by compensatory lengthening of the vowel. Examples are well documented for Old Saxon (Holthausen 1900: 68-69; Krogh 1996: 230-31), Old English (Campbell 1959: 47; Hogg 1992: 57), and Old Frisian (Boutkan and Siebinga 2005). Gothic (Go) and Old High German (OHG) did not undergo this change; thus, they remain true to the original Proto-Germanic forms. Examples of the Ingveonic innovation are given in (5).

[5] Ingveonic refers to a stage in the history of West Germanic. Old English, Old Saxon and Old Frisian are the daughter languages of Ingveonic.

(5) Examples of Ingveonic Nasal Deletion with Compensatory Lengthening[6]

	Gothic	Old Saxon	Old English	Old Frisian
'five'	*fimf*	*fīf*	*fīf*	*fīf*
'known'	*kunþs*	*kūth*	*cūþ*	*kūth*
'us'	*uns(is)*	*ūs*	*ūs*	*ūs*
'mouth'	*munþs*	*mūth*	*mūþ*	*mund*, mūþ*
'sense'	*gasinþa*	*gesīth*	*gesīða*	*sīth*
'tooth'	*tunþus*	*tand**	*tōþ*	*tōþ*

In (5), the canonical pattern of Ingveonic nasal deletion with compensatory lengthening can be seen. Old Saxon (OS), Old English (OE) and Old Frisian (OF) have a long vowel followed by a fricative, whereas Gothic retains the original VnF-sequence from Germanic. Thus, Germanic **fimf* > *fimf* in Gothic, but **fimf* > *fīf* in the Ingveonic group.

Because the sound change unconditionally affected all VnF-sequences, we should not expect to find any such sequences which cannot be explained by later diachronic change, and which have existed since Ingveonic times[7]. Yet, the word *kanst* 'you can,' shows an instance where the nasal is not deleted[8]. The complete Ingveonic paradigm of *kunan* 'to be able to' in (6) is compared with an incorrect model of the paradigm in (7). We should expect the incorrect paradigm given normal application of Ingveonic nasal deletion and compensatory lengthening. In the incorrect model, *kāst* is given as the second person singular form.

(6) The Actual Ingveonic Paradigm

kann	'know.1st.sg'	*kunan*	'know.1st.pl'
kanst	'know.2nd.sg'	*kunan*	'know.2nd.pl'
kann	'know.3rd.sg'	*kunan*	'know.3rd.pl'

(7) Paradigm Given Normal Application

kann	'know.1st.sg'	*kunan*	'know.1st.pl'
kāst	'know.2nd.sg'	*kunan*	'know.2nd.pl'
kann	'know.3rd.sg'	*kunan*	'know.3rd.pl'

[6] In the table in (5) the macron is used to denote a long vowel. Also note that orthographic <þ> and <th> both represent phonetic [θ]. An asterisk is used to indicate words where /n/ reentered through borrowing from Middle Low German and Middle Dutch.

[7] Systematic counterexamples to this generalization result from later syncope of a vowel between a nasal and a fricative. These counterexamples do not undergo compensatory lengthening and nasal deletion because they derive historically from VnVF sequences. For example, OS *hōnða* 'dishonor,' *minson,* 'diminish' < **hōnida* and **minnisōn (Holthausen 1900: 68)*. Apparent counterexamples may also come from later loan words OS, OF *tins* 'tribute, tax' < Latin *cēnsus*. I also assume that the Old Saxon and Old Frisian word *anst* 'grace,' was borrowed back into the languages from OHG (which regularly did not have nasal deletion and compensatory lengthening). Borrowing from OHG is quite common in the extant OS corpora. Moreover, *anst* is attested in OHG (Braune 1967: 201). Unlike the other Ingveonic languages, Old English seems to have maintained the original word; thus, we find the expected form *ēst*.

[8] Note that Ingveonic *kanst* derives from Germanic **kannt*, the latter form being attested in ON and Gothic. I assume, however, that the morphological change, whereby –t > –st, occurred between stages of NWGmc. and WGmc. The reconstructive evidence for this claim is that *all* the WGmc languages– the Ingveonic Group as well as OHG– have the form *kanst*. While there is no written record to go on (and therefore there can be no absolute certainty that the change from –t > –st necessarily took place at this early date), any argument that posits a later development of the morphological change –t > –st crucially assumes that –t > –st on *four* independent occasions in the WGmc. group. Namely, it would have to have transpired independently in OE, OS, OF and OHG alike. I assume that this quadruplication is exceedingly unlikely and that the change –t > –st, accepting some very small margin of error, took place long before the advent of the Ingveonic group.

In (6), all the cells of the paradigm have a nasal consonant. However, the second person singular cell contains a VnF sequence, which ought regularly to have undergone nasal deletion and compensatory lengthening. If the second person cell had undergone this change, then the paradigm would be as in (7). That is, (7) reflects normal application of the historical change. At the same time, the second person singular cell in (7) becomes anomalous with respect to the nasal feature, which appears in every other cell of the paradigm.

The aberrant example, *kanst*, in (6) can be explained if we assume that something is exerting an influential force on the word *kanst* so that it maintains the nasal feature common to the other members of the paradigm. The logical source of the influence is the base. The reasoning behind this statement is that the base is the only thing that all the forms have in common with each other. Thus, it can be said that the base of the paradigm, [kan] (which is both an existing output form and a form that is compositionally related to *kanst*), establishes an output that interacts with the evaluation function. As a result, an opaque surface form results, namely, a word with a nasal, where a nasal is locally forbidden.

Formalizing opacity involves some highly ranked constraint that trumps the markedness and faithfulness constraints governing normal application of a sound change. The constraints for such an analysis are given in (8).

(8) Constraints for a Base-Identity Analysis

a. BASEIDENT(αnas)
"For some segment α in the base, and some segment β in the corresponding output, let there be agreement with respect to [αnas]"
b. *NF
"Nasals may not precede fricatives"
c. IDENT-IO(αnas)
"For some segment α in the input, and some segment β in the corresponding output form, let there be agreement with respect to [αnas]"

The BASEIDENT(αnas) constraint is an output-output constraint. It assigns a violation mark whenever compositionally related words are not uniform with respect to a nasal segment. *NF is a markedness constraint stating that sequences of nasals followed by fricatives are ungrammatical. Finally, IDENT-IO(αnas) is a faithfulness constraint which requires a nasal segment in the input to be faithfully represented in the output.

The data in (5) demonstrate that *NF is crucially ranked above IDENT-IO(αnas). If the constraints appeared in the reverse order, then nasals would not be allowed to delete. The tableau in (9) formalizes the crucial ranking for the canonical pattern of nasal deletion and compensatory lengthening.

(9) Crucial Ranking of *NF and IDENT-IO(αnas)[9]

Input: / fimf /	*NF	IDENT-IO(αnas)
a. fimf	*!	
b. ☞ fĩf		*

The only way for the correct output form to be predicted is if the constraints are as they appear in (9). Reversing the ranking will cause the archaic form to incorrectly be selected. Thus, Ingveonic languages must have the crucial ranking of *NF >> IDENT-IO(αnas), while the Germanic languages[10] that retained the nasal, namely Go and OHG, have the ranking IDENT-IO(αnas) >> *NF. However, the ranking of constraints in (9) fails to explain the opaque output of words like *kanst*. This made explicit in (10).

(10) The Problem of *kanst*

Input: / kanst /	*NF	IDENT-IO(αnas)
a. kanst	*!	
b. ☹ kāst		*

The candidate *kāst* is an incorrect winner because the rankings for the canonical pattern fail to capture instances where some output form exerts an influential force on the evaluation of a word. In other words, the constraints in (10) are unable to predict an opacity effect. To deal with this limitation, a base-identity constraint is included along with a base. Faithfulness to the base serves as an argument why the intended winner loses.

(11) Motivating Opacity with a Base

Input: / kanst / Base: [kan]	BASEIDENT(αnas)	*NF	IDENT-IO(αnas)
a. ☞ kanst		*	
b. kāst	*!		*

[9]Accounting for compensatory lengthening in OT has some limitations. One cannot necessarily assume that the lengthening occurs because the mora from the deleted nasal is reassigned to the preceding vowel. This notion would require a mora faithfulness constraint, MAX-μ. However, it is controversial to assume that a mora can be assigned to consonantal segments in the input. Such segments are generally assumed not to be parsed into prosodic constituents. Given this assumption, the necessary identity constraint, MAX-μ, is theoretically impermissible; no mora can be assigned to the /m/ or /f/ of the input /fimf/. This problem will not be taken up in further detail, as it is peripheral to the goals of this paper. However, I do not consider [fif] as a possible candidate because of this limitation.

[10] North Germanic languages are not considered here because the VnF sequences in these languages underwent different changes. These developments make North Germanic languages less representative of the original Germanic forms. Generally, the nasal was kept in VnF-sequences, except before /s/. Thus, Go *fimf* 'five' corresponds to Old Norse *fimm*, while Go *uns(is)* corresponds to ON *oss* (cf. Robinson 1992: 88, 122).

In (11), it can be seen that the demotion of the BASEIDENT(αnas) constraint would involve the selection of the wrong candidate. Thus, that constraint must be ranked above the markedness constraint, *NF.

The tableau in (11) implies that the high-ranked sensitivity to output locally prevented a diachronic change from achieving fruition. In other words, the winning candidate *kanst* 'you can' reflects an archaism. In the following section, I will argue that social forces act very similarly to the base in that they often result in the resistance to a historical change that causes an antiquated form to be contextually appropriate. For example, Labov (2001: 319) has observed, "Almost all the sound changes in progress studied here and elsewhere involve gender differentiation," with females leading over men in the use of innovative forms (Labov 2001: 366-67; Cheshire 2002: 425-26). In section 4, I argue that the details of such facts are not incompatible with phonological theory, because output-output models offer the crucial window to the outside world that is missing from earlier models of variation in OT.

4. Output-Output Constraints as a Link to the Outside World: Vestige Theory

I suggested above that variation models in OT are deficient because they fail to establish a link to an individual's social context. In this section, I argue that the base-identity model of OT, which has traditionally been used to capture phenomena like paradigm uniformity, can be adapted as a means to incorporate sociolinguistic data into phonological theory. Specifically, I will assume that individuals can be seen as members of a social paradigm much the same way that one cell of an inflectional verb paradigm patterns with other cells of the paradigm to which it belongs.

Starting with the Base-Identity model, the paragraphs below form a derivation for what I referred to earlier as Vestige Theory. Vestige Theory assumes that all historical change results from constraint demotion (to be elaborated in section 5). In the demotion of any given constraint– *markedness and faithfulness constraints alike*– a vestige of that constraint is left behind in its original position.

The "true" demoted constraint and its vestige are assumed to be different from each other. The "true" constraint is a conventional input-output constraint, while the "vestige" is an output-output constraint. The vestige constraint is governed by social norms, which is to say that certain demographics appeal to the vestige constraint more than others. Demographics with a high appeal to the vestige constraint will use an antiquated form more than demographics with a low appeal to the vestige constraint.

Vestige Theory enables us to predict the existence of variation in OT. What it does not do, however, is formalize the exact nature of that variation. This is appropriate because variation is different from culture to culture, and we must therefore rely on the work of sociolinguists to describe the social factors that cause an appeal to the vestige constraint.

To begin, let us take the r-full, "Dialect A" analysis from (1), predicted by the rankings in (3).

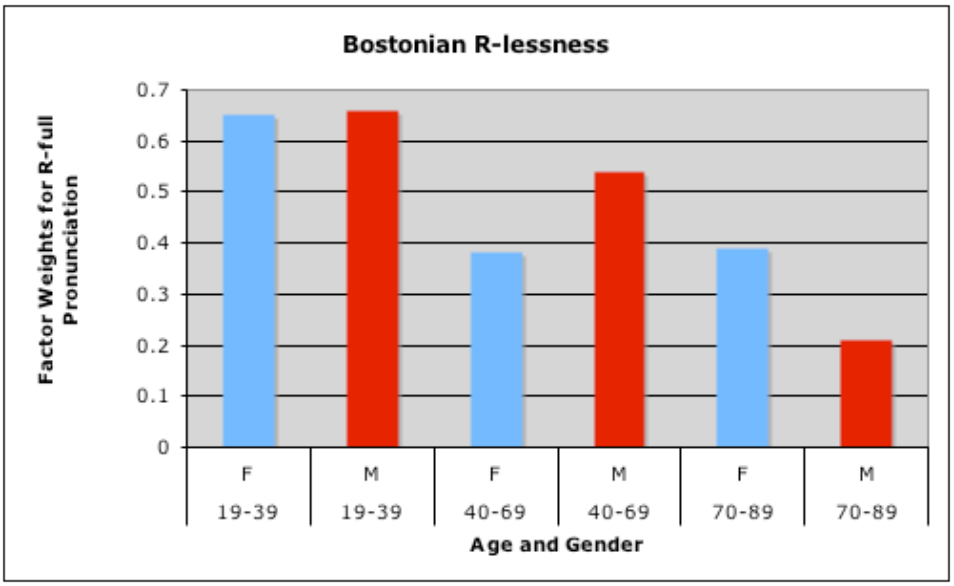
(12)

Input: / homər /	FAITH	*CODA
a. ☞ homer		*
b. home<r>	*	

A final rhotic is predicted to win in the tableau because FAITH to the rhotic segment of the input is crucially ranked above *CODA[11]. Were the ranking reversed, then the non-rhotic form would be predicted. The Tableau in (12) implies a categorical use of r-full forms. By the same token, switching the constraint ranking would predict a categorical employment of r-less pronunciations.

Contrary to these predictions, however, data from a Bostonian speech community indicate that neither extreme is ever realized. That is, the predictions in (12) cannot be reconciled with the data from (4), repeated in (13) for convenience:

(13)



The fact that older speakers tend to be r-less in their pronunciation while younger speakers tend to be r-full is entirely consistent with change in progress: older speakers use the form that was more common in older times, younger speakers use the form that is more common presently. This distribution implies that r-full pronunciation will continue to become more common (assuming no change in conditions). Similarly, we can project historically; 100 years ago and earlier, an even greater amount of r-lessness was extant. This is consistent with historical accounts of r-lessness; see, for example, Wolfram and Schilling-Estes (1998: 92-99).

Focusing on the older male speech (70-89) of this community, one observes a strong preference for the old-fashioned, non-rhotic pronunciation. Yet, it cannot be ignored, that these same speakers produce r-full utterances with a notable frequency. Below I argue that this exceptional behavior is similar to the exceptional behavior of Ingveonic *kanst* 'you know.'

Recall from above, that faithfulness to the base of the verbal paradigm caused the Ingveonic languages to resist historical change against VnF-Sequences. Because of faithfulness to other output, namely a base, the historical form *kanst* 'you know' was preserved. Tableau (17) repeats Tableau (14).

[11] The ONSET constraint discussed in (3) is superfluous to the discussion below because r-epenthesis does not occur in the particular dialect at hand.

(14)

Input: / kanst / Base: [kan]	BASEIDENT(αnas)	*NF	IDENT-IO(αnas)
a. ☞ kanst		*	
b. kāst	*!		*

The influence of the base results in the preservation of an archaic form or a resistance to language change. This resistance to change is similar to the tendency of the older Bostonian males in (13) to use the older, r-less pronunciation. Thus, one way to model the gender difference is to indicate that an older Bostonian male speaker's demographic exerts a pressure on him to conform to the linguistic norms of that social group. This pressure can be modeled analogously to the force (i.e. the uniformity effect) exerted on a grammatical cell because of that cell's membership to a grammatical paradigm. However, this proposal, represented below in (18), is only a stepping-stone. It will be rejected and modified in the paragraphs below.

(15) 'Homer' in the Speech of Older Bostonian Males (to be revised)

Input: / homər / Older Male: [homə]	OLDERMALEIDENT	FAITH	*CODA
a. homer	*!		*
b. ☞ home<r>		*	

In (15), the words "Older Male" below the input indicate the output used by the social paradigm at hand. That is to say, the form [homə] represents the output that older male speakers use: when an individual is in a social environment where he is surrounded by this output, the individual may conform to the norms of the social paradigm. The individual's decision to do so is represented by the constraint OLDERMALEIDENT. Whenever that individual evokes camaraderie with the social paradigm, the non-rhotic form surfaces. If, by contrast, the older male speaker were addressing some other individual, for example a younger university professor, the older male speaker would no longer feel the pressure to conform to his social paradigm. The OLDERMALEIDENT constraint becomes contextually inappropriate and is consciously avoided. In consequence, a rhotic form surfaces as the ranking FAITH >> *CODA predicts.

Some immediate revision must be made to the proposal advanced by the tableau in (15). The specific issue with such a proposal is that constraints, which are part of a cognitive *linguistic* representation cannot directly integrate *non-linguistic* factors.

To address this problem it is important to examine why the model in (15) can potentially work. It is because the output of the older men [homə] is the same as what the output would be if *CODA were highly ranked. It is as though the same constraint appears in two positions within the hierarchy:

(16) 'Homer' in the Speech of Older Bostonian Males (to be revised)

Input: / homər / Older Male: [homə]	*$CODA_1$	FAITH	*$CODA_2$
a. homer	*!		*
b. ☞ home<r>		*	

The tableau in (16) makes the same predictions as the one in (15). The difference between the two is that tableau (15) uses an ad hoc "social constraint," which implies that demographic behavior is *directly* part of the universal set of constraints. The tableau in (16), by contrast, gives demographics an *indirect* involvement with the universal set of constraints. That is, in the former case, the demographic is itself a universal constraint, while in the latter case, the demographic only influences the use of a universal constraint.

The difference between the two CODA constraints is that *CODA$_1$ is an output-output constraint that is evoked by an affinity to the output forms used in the individual's social environment, while *CODA$_2$ is an ordinary markedness constraint. Older male speakers therefore employ the *CODA$_1$ constraint because this is the output they have heard among others in their (Older Male) demographic. They align their speech to be consistent with the speech of their social paradigm, as it were.

What is striking about the tableau in (16), however, is that it seems to incorporate both the modern-day r-full constraint ranking as well as the r-less constraint ranking of the olden days. The two rankings are given in (17).

(17) Newer r-full ranking: FAITH >> *CODA$_2$
Older r-less ranking: *CODA$_1$>> FAITH

We can describe the change from r-less to r-full pronunciation as a demotion of the *CODA constraint:

(18)

Input: / homər / Older Male: [homə]	*CODA	FAITH	
a. homer	*!		
b. ☞ home<r>		*	

The arrow in (18) represents a historical process of constraint demotion. By comparing the tableau in (16) with the one in (18), we see that *CODA$_1$ reflects the historical position of *CODA from the time when r-less pronunciation was most common. *CODA$_2$ marks the spot in the contemporary hierarchy to where *CODA was moved. That is, that *CODA$_1$ in (16) is actually a *vestige* of *CODA from (18). The final representation for this is given in (19).

(19) 'Homer' in the Speech of Older Bostonian Males

Input: / homər / Output: [homə]	VESTIGE(*CODA)	FAITH	*CODA
a. homer	*!		*
b. ☞ home<r>		*	

The tableau in (19) suggests that the new constraint ranking (i.e. FAITH >> *CODA) is, in fact, present in older male speech. However, this demographic's appeal to VESTIGE(*CODA) from the original situation of the *CODA constraint evokes the socially contextualized surfacing of the

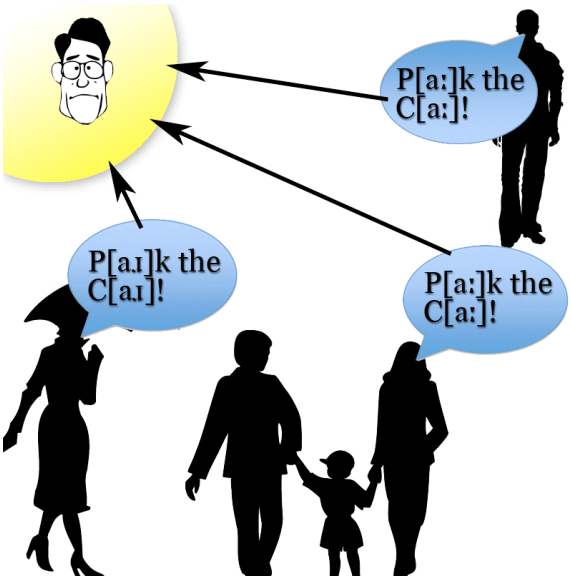
archaic form. When the older males do not appeal to the *vestige* constraint, then they use the r-full pronunciation predicted by FAITH >> *CODA.

A *vestige* is materially different from the constraint it shadows. Specifically, a *vestige* is a kind of output-output constraint, while the constraint it shadows is a "normal" markedness (or, as we will see below, a "normal" faithfulness) constraint. As an output-output constraint a *vestige* represents a *possibility* to articulate an archaic form. This possibility is licensed by the individual's constraint ranking, but importantly there is *nothing* in the individual's constraint ranking that indicates what specific factors condition variation. In other words, phonologists cannot perform the work of sociolinguists in OT.

The shading in (19) is used to help conceptualize the vestige constraint. The vestige constraint functions a bit like an old fluorescent bulb that blinks on and off. This blinking on and off corresponds to the use and non-use of archaic forms within a language community. The older Bostonian males from (19), use archaic forms relatively frequently and therefore the vestige constraint blinks with higher frequency. Conversely, younger speakers use archaic forms less regularly and therefore the blinking of the vestige constraint (i.e. the use of the archaic form within that social group) occurs infrequently.

Note that the vestige constraint is assumed to always be present in an individual's grammar, and that it is the variability within the community, the flashing use and non-use of archaic forms, that determines the rate of the vestige constraint's flickering. This is illustrated in (20).

(20) Saying "Park the car!" in a Bostonian Linguistic Community[12]



The illustration in (20) models a linguistic community with variable output. Some of the members of the community have non-rhotic pronunciation, others have rhotic pronunciation. This variable output is uttered to an individual, as represented by the arrows. In this model

[12] The diagram in (20) uses free images from http://www.fundraw.com/clipart/ and http://www.silhouettesclipart.com.

linguistic community, the r-less pronunciation is used (i.e. it flashes about within the community) twice as frequently as r-full pronunciation. We therefore expect that the vestige constraint of the individual will flash with the same frequency that the utterances occur within the linguistic community.

Thus, the proposal here is not superior in its ability to describe all the social factors that condition variation. It cannot. The specific advantage of Vestige Theory to the other models of variation in OT is that it references the outside world whereby social conditioning factors are *licensed*. Moreover, Vestige Theory is consistent with the fact that variation is intimately intertwined with language change. The theory predicts that every instantiation of a historical change (i.e. a constraint demotion) is also the instantiation of sociolinguistic variation. In the following section, I discuss the implications that Vestige Theory has on OT.

5. The Implications of Vestige Theory on OT

There are a number of important assumptions behind Vestige Theory. They are listed in (21).

(21) a. Variation is the result of pressure to conform to a social group's output norms.
b. Language Change is always the result of constraint demotion.
c. Appeal to the vestige constraint is something we manipulate to a degree.

In section 5.2 I discuss these assumptions using data from Norwich English develarization (often called "g-dropping"), which is described in 5.1.

5.1 An Analysis of Develarization ("g-dropping") in Norwich English

The data in (22a) and (22b) are based on Trudgill's (1974) classic study. They represent all the words ending in *–ing* that went into his analysis. [ŋ] ~ [n] indicates that variation exists in these words. Sometimes the word is pronounced with a velar nasal, though more often it is pronounced as a coronal nasal (further details to be discussed below).

(22)

(a) develarizing verbs		(b) develarizing nouns	
doing	[ŋ] ~ [n]	evening	[ŋ] ~ [n]
walking	[ŋ] ~ [n]	morning	[ŋ] ~ [n]
laughing	[ŋ] ~ [n]		
playing	[ŋ] ~ [n]		
coming	[ŋ] ~ [n]		
going	[ŋ] ~ [n]		
drawing	[ŋ] ~ [n]		
sawing	[ŋ] ~ [n]		
soaring	[ŋ] ~ [n]		
helping	[ŋ] ~ [n]		
falling	[ŋ] ~ [n]		

Temporarily treating each variant as its own dialect, we can begin to piece together an explanation for these data in Optimality Theory. Accounting for the coronal nasal is fairly straightforward. Only two constraints will be necessary. These are given in (23).

(23) 1. *ŋ]$_{\sigma}$
"No final velar nasals at the right edge of an unstressed syllable[13]."
2. IDENT-IO(Place)
"The input and output agree with respect to place."

The first constraint is a contextual markedness constraint that militates against foot-final [ŋ]. The second constraint is faithfulness constraint that keeps a dorsal place of articulation from becoming coronal. The Tableaux in (24) predict that the coronal nasal will emerge when *ŋ]$_{\sigma}$ is high-ranked, while the velar nasal is retained when IDENT-IO(Place) has the highest rank[14].

(24) *Doing* in a Dialect with Velar Nasal Retention

Input: / duwiŋ /	IDENT-IO(Place)	*ŋ]$_{\sigma}$
a. ☞ duwiŋ		*
b. duwin	*!	

Doing in a Dialect with Develarization

Input: / duwiŋ /	*ŋ]$_{\sigma}$	IDENT-IO(Place)
a. duwiŋ	*!	
b. ☞ duwin		*

The data in (25) do not exhibit variation. These data stand as apparent counter-examples to the generalization captured by the tableaux in (24).

(25)

Verbs		Nouns		Adjectives	
sing	[ŋ]	king	[ŋ]	strong	[ŋ]
sang	[ŋ]	string	[ŋ]	young	[ŋ]
sung	[ŋ]			wrong	[ŋ]
ring	[ŋ]				
rang	[ŋ]				
rung	[ŋ]				

[13] This constraint is best characterized as a language specific one. Debate on the validity and necessity of language specific constraints in OT extends well beyond the purpose of this paper. The constraint is used here for convenience; better universal explanations are open to further research. An important assumption is that *linguistic* constraints that are language specific are materially different from *non-linguistic* constraints that are language specific. While ample evidence for the former has been argued in countless scholarly works, there are no serious proposals to include *non-linguistic* constraints in a formal model of OT.

[14] For space concerns, I am not considering the labial nasal candidate. I assume that the coronal place of articulation is the unmarked point of articulation in English. An analysis could therefore assess a double violation against the labial place of articulation. Further research might consider additional alternatives.

Tagliamonte (Tagliamonte 2004, cited in Foulkes and Dougherty(2007:64)) examines the similar York dialect and finds that part-of-speech is the most significant predictor of variation. Specifically, she argues that verbs are much more likely to exhibit coronalization of the velar nasal than nouns or adjectives. Note however, that the data in (22b) contradict this generalization. While it could be argued that develarization in these words is lexically conditioned, a closer look at the data offers a more straightforward explanation. The data in (25) are all monosyllabic. The velar nasal in these words therefore occurs in a stressed syllable. By contrast, the words in (22) have velar nasals that stand in unstressed syllables. The constraint *ŋ]$_{\sigma}$ is therefore superfluous with respect to the data in (26), resulting in the selection of the velar nasal:

(26) *Sing* in a Dialect with Velar Nasal Retention

Input: / siŋ /	*ŋ]$_{\sigma}$	IDENT-IO(Place)
a. ☞ siŋ		
b. sin		*!

Sing in a Dialect with Develarization

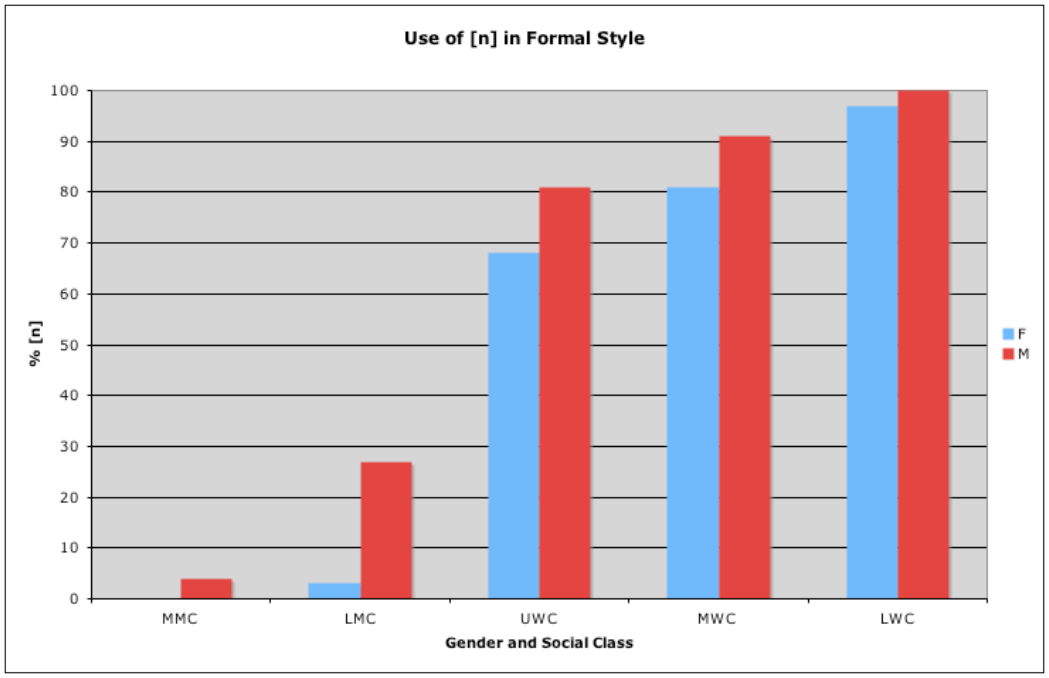
Input: / siŋ /	IDENT-IO(Place)	*ŋ]$_{\sigma}$
a. ☞ siŋ		
b. sin	*!	

The tableau in (26) predicts that dialects with develarization will not change the place of articulation in words like *sing* because the [ŋ] of the output is at the right edge of a stressed syllable.

5.2 Variation in Norwich Develarization and the Assumptions of Vestige Theory

The tableaux in (24) predict the two nasal variants [ŋ] and [n]. The assumption of these tableaux is that speakers will categorically use one form or the other depending on the individual's constraint ranking. The problem with these tableaux, however, is that they do not take into account the sociolinguistic facts given in (27).

(27)



In (27), we see the use of [n] and [ŋ] in the formal style of the Norwich community according to different demographics. The middle middle class (MMC) females, far left, exhibit categorical employment of [ŋ] while the lower working class (LWC) males, far right, have unconditional develarization. The former demographic seems to have a high ranking of IDENT-IO(Place), the latter *ŋ]$_{\sigma}$. But both rankings would be necessary to account for the data in the remaining demographics.

1. Variation is the result of pressure to conform to a social group's output norms. Vestige Theory claims that the distribution in (27) is the result of each demographic's appeal to the vestige constraint. This appeal is not a phonological process, but rather a sociolinguistic phenomenon stemming from a social norm. What can be modeled phonologically, however, is the vestige itself, which is nothing more than the phantom of a demoted constraint:

(28) *Doing* in a Dialect with Velar Nasal Retention

Input: / duwiŋ /		Vestige(IDENT-IO(Place))	*ŋ]$_{\sigma}$	IDENT-IO(Place)
a.	duwiŋ		*!	
b. ☞	duwin	*!		*

The tableau in (28) assumes that the coronal nasal pronunciation is the innovative one. This assumption is based simply on what we know of the etymology of words like the one in (28): *ŋg > ŋ > n (cf. Wells 1982); the voiced velar stop of an original [ŋg] sequence was deleted, and the remaining nasal segment, [ŋ], was subsequently develarized to [n]. Once the vestige is established the tableau can be read in conjunction with the data in (27) to derive a complete phonological and sociolinguistic picture. That is that the data in (27) give facts about different demographics' appeal to the vestige constraint while the tableau in (28) gives an argument as to the cause or origin of the vestige.

2. Language Change is always the result of constraint demotion. Note in (28) that the faithfulness constraint is being demoted. This may seem a little counterintuitive at first because language change is often characterized by markedness reduction (Kroch 1978; Vennemann 1993). In OT it may seem more intuitive to model markedness reduction as a promotion of markedness constraints, a notion which has been expressly argued by Cho (1998). However, OT only stipulates that this kind of language change results when some markedness constraint dominates some faithfulness constraint. There are two ways that this can occur. The first is via a promotion of the markedness constraint above the faithfulness constraint (e.g. Cho 1998). The second is via a demotion of the faithfulness constraint below the markedness constraint. Vestige Theory necessarily assumes the latter[15]; if the markedness were promoted, then the vestige would be ranked below the constraint it shadows and there would be no way for a social group to make an appeal to the vestige constraint. The tableau in (29) demonstrates the problem with viewing language change as a promotion of a markedness constraint.

(29) *Doing* in a Dialect with Velar Nasal Retention

Input: / duwiŋ /	*ŋ]$_{\sigma}$	IDENT-IO(Place)	Vestige (*ŋ]$_{\sigma}$)
a. duwiŋ	*!		*
b. ☞ duwin		*!	

Because the markedness and faithfulness constraints are ranked above the vestige constraint, the vestige constraint cannot have any affect on the selection of the winning candidate. In this case, we would expect variation *never* to occur. Since variation surrounds all language change, Vestige Theory predicts that language change is always the result of constraint demotion. Therefore local decreases in markedness are seen as a reduction in faithfulness rather than an increase in well-formedness.

Vestige Theory predicts that sound change is not always a local improvement to markedness. It also predicts that the historical development of more marked structures may occur. Where decreased markedness involves the *demotion of faithfulness* constraints, increased markedness can simply be expressed as the *demotion of markedness* constraints. This is what we saw in (22), repeated in (30) for convenience.

(30) 'Homer' in the Speech of Older Bostonian Males

Input: / homər / Output: [homə]	VESTIGE(*CODA)	FAITH	*CODA
a. homer	*!		*
b. ☞ home<r>		*	

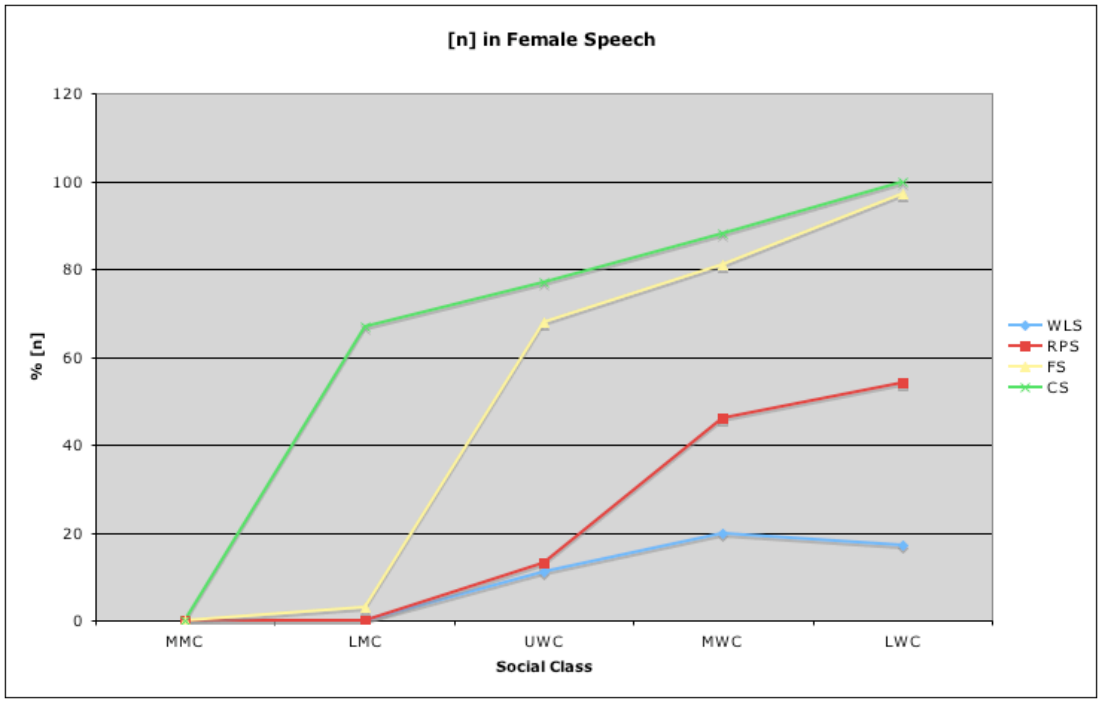
The less marked non-rhotic pronunciation is abandoned in favor of a more marked r-full pronunciation via the demotion of the markedness constraint *CODA. According to Vestige Theory, what all sound changes have in common– whether a local improvement or regression to

[15] Note that constraint demotion is a process argued for by Tesar and Smolensky (2000) for reasons independent of variation. This is discussed in McCarthy (2008).

markedness– is that they involve a constraint demotions and that the vestiges of these demoted constraints enable variation between an older form and a contemporary one.

3. Vestige constraints are something we can manipulate to a degree. In (31) more data are given from Trudgill (1974).

(31)



In Norwich English, female speakers from every social class generally exhibit different pronunciations depending on how much attention they are giving their pronunciation. The lowest line shows that, when reading word lists (WLS), the females use [ŋ] 80 percent of the time or more. The next three lines up from the bottom respectively depict the forms that are elicited while reading a passage (RPS), using formal speech (FS), and using casual speech (CS). With the exception of the middle middle class females, the more cautious the speech, the fewer instances of develarization. These data suggest that formality factors motivate an individual to appeal to certain forms, a tendency that has been uncontroversially interpreted in countless other sociolinguistic studies (cf. Schilling-Estes 2002: 375-96; Chambers 2006: 4). The fact that stylistic variation exists is a tremendous shortcoming for traditional OT models because even if they can define an individual's grammar to a very specific demographic, these models still would have the task of explaining the variant pronunciation of the individual himself.

The pronunciation of an individual can change with his lifetime as well (Sankoff and Blondeau 2007). This occurs as the result of strong social pressures. In Vestige Theory, it is these output pressures that motivate the individual to appeal more or less frequently to a vestige constraint.

6. Previous Models of Variation in OT

In this section, I examine previous models of variation in OT. I argue that a Vestige Theory account of variation is superior to the previous proposals because the alternatives do not have a link to the outside world. The Bostonian data from (1) are repeated in (32) for convenience:

(32)

Dialect A	Dialect B	Dialect C
Wanda left	Wanda left	Wanda left
Homer left	Home<r> left	Home<r> left
Wanda arrived	Wanda arrived	Wanda[r] arrived
Homer arrived	Homer arrived	Homer arrived

The reader will recall from (3) that the constraint rankings in (33) can be used to distinguished one dialect from the other:

(33)

	Constraint Ranking
Dialect A:	FAITH >> *CODA >> ONSET
Dialect B:	*CODA >> FAITH >> ONSET
Dialect C:	*CODA >> ONSET >> FAITH

If one assumes that speakers of Bostonian English have all three dialect grammars and that the possession of multiple grammars is responsible for the variant surface forms, there are two theoretical problems that arise. The first is an issue of elegance. Dialects have hundreds of variants in them. Variation in Bostonian English is certainly not restricted to the rhotic phenomena described by the data in (32). There are scores of other characteristics that circumscribe the Bostonian English dialect. Note therefore, that any additional variant necessitates additional grammars. In fact, the number of needed grammars multiplies exponentially with the inclusion of every new dialect variant. That is, if another variant were explained by the constraint rankings X, Y and Z, one would have to talk about Dialect A_x, Dialect A_y, Dialect A_z, Dialect B_x, Dialect B_y, etc. Millions of grammars would be required in order to describe the linguistic behavior of any given individual.

Beyond this, were the millions of grammars considered a viable analysis, there is *still* no link to the external world that explains why those grammars work to produce different output when the individual is in one social context as opposed to another.

Some of these problems are addressed in Anttila and Cho (1998: 36), who argue that the limitations of traditional OT with respect to variation stem from an assumption that all constraints are a series of ordered pairs within a grammar's hierarchy. As ordered pairs, these constraints possess four properties, which are given below.

(34) a. *Irreflexivity*
'No constraint can be ranked above or below itself'

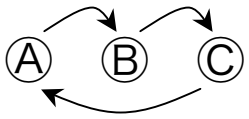
b. *Asymmetry*
'If X is ranked above Y, it cannot be ranked below Y'

c. *Transitivity*
'If X is ranked above Y, and Y is ranked above Z, then X is ranked above Z"

d. *Connectedness*
'Crucial rankings exist between constraints'

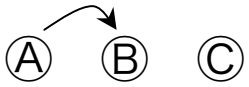
Anttila and Cho (1998) argue that variation is a predictable outcome whenever the connectedness property does not hold among a group of constraints. In other words, given constraints A, B and C, no variation should occur if these constraints remain connected (i.e. crucially ranked), for example, A >> B, B >> C, A >> C. In this case, every constraint is ranked with respect to every other constraint:

(35)



The arrows in (35) represent that two constraints are *connected*. That is, some crucial ranking exists between them. For instance, the arrow between constraints A and B could represent the crucial ranking A >> B (note however, that it could also represent B >> A). Variation will occur, as Anttila and Cho (1998) claim, whenever some of the connectedness relationships are missing. For example:

(36)



In (36), we expect some free variation since some of the constraints are not connected. Specifically, the diagram licenses three pronunciations. Each variant is defined by the constraint relationship between A and B, for example, A >> B. Because C is not connected, however, it is free to be ranked above A, between A and B, or below B. Thus, three constraint rankings are generated, namely C >> A >> B, A >> C >> B or A >> B >> C. Each of these rankings is presupposed to be a grammar that an individual uses, which is to say that variation is an operation of constraint shuffling that results from breeches in connectedness.

A stark advantage of this model is that is dispenses with the need to have superfluous grammars. All variation within a language can be theoretically accounted for with one constraint hierarchy. However, the model makes the prediction that all variants should be used in equal proportions. If there are three variants, then each should have a 33.33% likelihood of occurring. If there are only two variants, then each should occur with a 50% probability. Such is almost never the case, however. Multiple variants generally are employed with different probabilities.

Additionally, this model fails to have a link to the external world. It cannot explain why an individual exhibits one linguistic behavior in one context and another linguistic behavior in a separate context.

Boersma and Hayes (2001) have proposed an idea that is similar to Anttila and Cho (1998). Rather than dispensing with any of the traditional ordered pair assumptions given in (34), the authors propose that variation is the result of degrees of fuzziness among these properties. This fuzziness enables Boersma and Hayes (2001) to suggest that constraints overlap and that variation is the consequence of overlapping constraints. The Boersma and Hayes (2001) model of variation has one major advantage over Anttila and Cho (1998). The degree to which one variant is chosen over another form can be said to reflect the extent to which two constraints overlap. Thus, free variation indicates two completely overlapping constraints, non-overlapping constraints designate the absence of variation, and partially overlapping constraints denote a preference for one form over another.

One criticism of Boersma and Hayes (2001) is discussed by Auger (2001: 292-94), who observes that variation in the Boersma and Hayes (2001) model necessarily assumes the contiguity of the constraints involved. Future research must determine whether variation is possible among non-abutting constraints. Additional questions are brought up in Anttila (2002).

The larger issue that all the models face is that "they do not identify the [external] factors that condition… variation" (Irwin and Nagy 2007: 135). The fact of the matter is that variation data pose numerous challenges to the OT framework (cf. Anttila 2002 for an excellent review and critique of proposals). Although Anttila and Cho (1998) and Boersma and Hayes (2001) have developed analyses that license the selection of multiple candidates, there is nothing in their phonological frameworks that can effectively license external influence. For this reason, there remain aspects of variation that simply are not– and, crucially, cannot be– captured by these models.

Boersma and Hayes (2001) give some further attention to stylistic variation in Appendix C of their article. They argue that their model could have a link to the outside world by adding variables to the ranking algorithm. The authors suggest including a variable called *styleSensitivity*. Along these lines, it would be possible for Boersma and Hayes (2001) to improve their model by adding a full range of sociolinguistic variables. These variables would include *genderSensitivity, ageSensitivity, classSensitivity,* etc.

This analysis will ultimately fail because the equation involved in the computation of the ranking algorithm remains a language internal feature that is blind to language external data. This becomes transparent when we consider more than one variable feature within a language community. Say, for example, there were a dialect with variable r-less and r-full pronunciation (e.g. floo<r> ~ floo[r] 'floor'). In addition to the variant rhotic pronunciation, suppose the dialect also had variable velar and coronal nasals (e.g. goi[n] ~ goi[ŋ] 'going'). It is nearly impossible that these two variable features would have identical distributions. That is, r-full pronunciations might occur 45% of the time (and r-less pronunciations 55% of the time) while velar nasals might be articulated 80% of the time (and develarized 20% of the time).

The *Sensitivity* variables (*styleSensitivity*, *genderSensitivity, ageSensitivity, classSensitivity,* etc.) seem to predict that the same amount of overlap from the ranking algorithm should apply to all affected constraints. However. the myriad variable features of any given dialect will seldom, if ever, have the same degree of constraint overlap. This is simply because different variables play a role for different variants in different ways. In other words, a good theory of variation needs to account for a situation where, for instance, r-less pronunciation

is associated with a lower income bracket while velar nasal pronunciation is associated with female speech.

The only way around these problems is to say that all the *Sensitivity* variables are encoded into every single constraint in the hierarchy. However, this solution dispenses with the notion that constrains are universal and we are left with a proposal that simply does not work.

Vestige Theory eliminates these problems because variation remains entirely external. The output-output constraint is the only language internal structure that is open to the external world. This structure, I have argued, is sufficient to *license* the fact that intricacies of our surrounding environment affect the way we speak at any moment in time.

7. Prospectus

In this article, I have proposed an anti-modular view of the external factors that condition variation. The vestige left behind by a demoted constraint acts as an output-output constraint whose application is governed by the social norms of demographics within a speech community.

There is much left to the work of future research. Vestige Theory implies a model of diachronic change for Optimality Theory. On the one hand, this is desirable since synchronic alternations stem from historical developments. In the examples discussed throughout this article, the input has been minimally important. However, it is clear that some lexical restructuring must occur, otherwise "all the Indo-European languages would necessarily have the same input lexicon as English…" (Reiss 2003: 143). An exact formalization of how and when lexical restructuring occurs has not been addressed; though it is an important question for future research to assay.

One argument against Vestige Theory might be called the "over-prediction" problem. This refers to the observation that the rankings in (37) formally appear to be the same.

(37) The "Over-Prediction" Problem
a. VESTIGE(MARKEDNESS) >> FAITHFULNESS >> MARKEDNESS
b. VESTIGE(FAITHFULNESS) >> MARKEDNESS >> FAITHFULNESS

Both the constraint ranking in (37a) as well as the one in (37b) predict variation between two forms. In other words, variation between some form A and some form B can be the result of demoted markedness (as in (19)) or the result of demoted faithfulness (as in (28)). If both explanations are possible, how can we synchronically distinguish MARKEDNESS demotion from FAITHFULNESS demotion? The answer to this question stems from the fact that Vestige Theory is a model of underapplication. That is to say that the application of certain sound changes does not take place in all demographics at the same rate. Demographics must be divided into the (change) applying population and the (change) underapplying population. In this view, the ranking in (37a) is fundamentally different from the ranking in (37b). In the former case, the underapplying population is maintaining the less marked pronunciation while the applying population is characterized by input faithfulness (FAITHFULNESS >> MARKEDNESS). This input faithfulness ranking (37a) is what unifies the linguistic community as a whole. The ranking in (37a) predicts, ceteris paribus, that there will be increasing input faithful pronunciations. By contrast, (37b) involves an underapplying population that maintains input faithfulness. However, the linguistic community as a whole is characterized by its favoring of less marked pronunciation

(MARKEDNESS >> FAITHFULNESS). If the ranking in (37b) remains unchanged, less marked output forms will increase in their use throughout the linguistic community.

The increase and decrease in use assumes that the underapplying population will eventually die out (quite literally) and that the applying population therefore increases. The increase of the applying population necessarily involves the decrease of the underapplication forms that are uttered within the linguistic community. It stands to reason that when the underapplying population is completely gone, younger generations would no longer learn the vestige constraint. I leave the terminal status of vestige constraints open to further research.

Vestige Theory dispenses with the Multiple Grammars Model of variation (cf. Anttila 2002) since these models claim that multiple rankings produce multiple forms. What these models are actually doing is enabling a constraint to be high-ranked in some instances and low-ranked in others. I have argued that the variant high-ranked constraint is actually the vestige of the lower-ranked constraint, and that this vestige, as an output-output constraint, flashes on and off according to the output norms of the social context in which an individual is situated.

The Multiple Grammars Model still has its applications, however. It is uncontroversial, claims Anttila (2002: 219), "that an individual can simultaneously possess several grammars… in the case of multilingualism. The question is whether multilingualism, multidialectalism, and an individual's ability to switch among styles, and registers… are fundamentally similar phenomena." While Vestige Theory suggests that these are not similar phenomena, some shades of grey can be found in languages like Yukaghir (as reported in Wardhaugh 2002: 316), an East Siberian language and Chamic Vietnamese (Blood 1967). In these languages gender based alternations have been observed.

"In Yukaghir… a male goes through the progression of /ts/, /tj/, and /čj/ and /dz/, /dj/, and /ǰj/" within his lifetime, using /ts/ and /dz/ as a child, /tj/ and /dj/ as an adult and /čj/ and /ǰj/ in old age. By contrast, females maintain /ts/ and /dz/ through adulthood, but do take on the same old-age pronunciations of men[16]. Thus, woman never pronounce /tj/ and /dj/.

In Chamic Vietnamese, Blood (1967: 21) observes that a substitution of [j] for [ɾ] "is especially noticeable in women's speech." Some examples are given below:

	Female Speech	Male Speech
'heavy'	tjaʔ	tɾaʔ
'day'	hjɛj	hɾɛj
'bedbug'	jah	ɾah

Vestige Theory can account for situations of gender variation like the ones above; however, a Vestige Theory account may rival a Multiple Grammar Model explanation. Future research is needed on the historical development of such phenomena. It does not seem to be known whether these situations generally last for centuries or for a few decades; the former scenario is, to my knowledge, unattested, and the latter has been observed in a dialect of Cretan Greek (Mansfield and Trudgill 1994) as part of a change in progress.

[16] Elena Maslova (personal communication), author of A Grammar of Kolyma Yukaghir (2003), was not able to confirm this pattern, and was skeptical about its authenticity. Wardhaugh does not provide a source for these data. Given that Yukaghir is currently moribund, these data may only hold true for a specific dialect at some unspecified period in time.

Vestige Theory contributes additional explanatory power to OT without doing much damage to the traditional assumptions of the model. The specific advantage of Vestige Theory is that it provides a possible explanation for noise in datasets. Some unexpected forms that have been preserved can be systematically eliminated from datasets: these data can be said to show the effects of a vestige constraint, i.e. they are data that specifically cannot factor into the argumentation of a synchronic constraint ranking, because they were licensed through an older one.

While the discussion above has examined noise from variation, Vestige Theory might also be productive to account for some of the noise that appears to be the result of frequency effects (cf. Phillips 1984). These kind of data often come up in contexts where OT, and in some cases, even the field of Phonology, are being expressly rejected. The proposal here suggests that there is still room for OT to operate despite the commonplace intricacies of linguistic data.

References

Anttila, A., 2002. Variation and Phonological Theory. In: Chambers, J. K., et al. (Eds.), The Handbook of Language Variation and Change, Blackwell Publishing, Malden, pp. 206-243.

Anttila, A. and Cho, Y., 1998. Variation and Change in Optimality Theory. Lingua 104, pp. 31-56

Auger, J., 2001. Phonological Variation and Optimality Theory: Evidence from Word-Initial Vowel Epenthesis in Picard. Language Variation and Change 13, pp. 253-303

Bell, A., 1991. Audience Accommodation in the Mass Media. In: Giles, H., et al. (Eds.), Contexts of Accommodation- Developments in Applied Sociolinguistics, Cambridge University Press, Cambridge, pp. 69-102.

Blood, D. L., 1967. Phonological Units in Cham. Anthropological Linguistics 9, pp. 15-31

Boersma, P. and Hayes, B., 2001. Empirical Tests of the Gradual Learning Algorithm. Linguistic Inquiry 32, pp. 45-86

Boutkan, D. and Siebinga, S. M., 2005. Old Frisian Etymological Dictionary. Koninklijke Brill NV, Brill.

Braune, W., 1967. Althochdeutsche Grammatik. Max Niemeyer Verlag, Tübingen.

Campbell, A., 1959. Old English Grammar. The Clarendon Press, Oxford.

Chambers, J. K., 2006. Sociolinguistic Theory: Linguistic Variation and its Social Significance. Blackwell Publishers, Malden.

Cheshire, J., 2002. Sex and Gender in Variationist Research. In: Chambers, J. K., et al. (Eds.), The Handbook of Language Variation and Change, Blackwell Publishing, Malden, pp. 423-443.

Cho, Y., 1998. Language Change as Reranking of Constraints. In: Hogg, R. M. and van Bergen, L. (Eds.), Historical Linguistics 1995, vol. 2. John Benjamins, Amsterdam, pp. 45-62.

Downing, L., et al. (Eds.), 2005. Paradigms in Phonological Theory. Oxford University Press, Oxford

Hay, J. and Sudbury, A., 2005. How Rhoticity Became /r/-Sandhi. Language 81, pp. 799-823

Hogg, R. M., 1992. A Grammar of Old English. Blackwell, Oxford.

Holthausen, F., 1900. Altsächsisches Elementarbuch. Carl Winter's Universitätsbuchhandlung, Heidelberg.

Irwin, P. and Nagy, N., 2007. Bostonians /r/ Speaking: A Quantitative Look at (R) in Boston. University of Pennsylvania Working Papers in Linguistics 13, pp. 135-147

Kroch, A., 1978. Toward a theory of social dialect variation. Language in Society 7, pp. 17-36

Krogh, S., 1996. Die Stellung des Altsächsischen im Rahmen der germanischen Sprachen. Vandenhoeck

& Ruprecht, Göttingen.
Labov, W., 1966. The Social Stratification of English in New York City. Washington, DC: Center for Applied Linguistics.
Labov, W., 1972. Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia.
Labov, W., 2001. Principles of Linguistic Change: Social Factors. Blackwell Publishers, Malden.
Mansfield, P. and Trudgill, P., 1994. A Sex-Specific Linguistic Feature in a European Dialect. Multilingua 13, pp. 381-386
Maslova, E., 2003. A grammar of Kolyma Yukaghir. Mouton de Gruyter, Berlin.
McCarthy, J. J., 1993. A Case of Surface Constraint Violation. Canadian Journal of Linguistics 38, pp. 169-195
McCarthy, J., 2008. Doing Optimality Theory: Applying Theory to Data. Blackwell Publishing, Malden.
Prince, A. and Smolensky, P., 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Rutgers University/University of Colorado at Boulder.
Philips, B., 1984. Word Frequency and the Actuation of Sound Change. Language 60, pp. 320-340
Reiss, C., 2003. Language Change without Constraint Reranking. In: Holt, E. (Ed.), Optimality Theory and Language Change, Kluwer Academic Publishers, Dordrecht, pp. 143-168.
Robinson, O. W., 1992. Old English and its Closest Relatives: A Survey of the Earliest Germanic Languages. Stanford University Press, Sanford.
Sankoff, G. and Blondeau, H. l. n., 2007. Language change across the lifespan: /r/ in Montreal French. Language 83, pp. 560-588
Schilling-Estes, N., 2002. Investigating Stylistic Variation. In: Chambers, J. K., et al. (Eds.), The Handbook of Language Variation and Change, Blackwell Publishing, Malden, pp. 375-401.
Tagliamonte, S. A., 2004. Somethi[ŋ]s goi[n] on! In: Gunnarsson, B. L., et al. (Eds.), Language Variation in Europe, Uppsala University Press, Uppsala, pp. 390-403.
Tagliamonte, S. A., 2006. Analyzing Sociolinguistic Variation. Cambridge University Press, Cambridge.
Tesar, B. and Smolensky, P., 2000. Learnability in Optimality Theory. MIT Press, Cambridge.
Trudgill, P., 1974. The Social Differentiation of English in Norwich. Cambridge University Press, Cambridge.
Vennemann, T., 1993. Language change as language improvement. In: Jones, C. (Ed.), Historical Linguistics: Problems and Perspectives, Longman, London, pp. 319-344.
Wardhaugh, R., 2002. An Introduction to Sociolinguistics. Blackwell Publishing, Malden.
Weinreich, U., et al., 1968. Empirical Foundations for a Theory of Language Change. In: Lehmann, W. P. and Malkiel, Y. (Eds.), Directions for Historical Linguistics: A Symposium, University of Texas Press, Austin, pp.
Wells, J. C., 1982. Accents of English. Cambridge University Press, Cambridge.
Wolfram, W. and Schilling-Estes, N., 1998. American English: Dialects and Variation. Blackwell Publishers, Malden.