Convergence is predicted by particular interlocutors, not speakers

Uriel Cohen Priva* Chelsea Sanker*

* Brown University, Department of Cognitive, Linguistics, and Psychological Sciences.
* Both authors contributed equally to this manuscript, and are listed alphabetically.

Corresponding author: Uriel Cohen Priva, Brown University

Abstract

Do characteristics of individuals affect how much they converge, such that some speakers consistently converge more than others? Are there aspects of conversations or interlocutors that elicit more or less convergence? One might expect variation in factors like attention to detail or sociability to produce individual tendencies in convergence, while factors like attention and arousal could produce conversation-specific or interlocutor-specific patterns, if certain individuals are more likely to inspire attention or excitement. We present data on convergence in the Switchboard corpus, compared across speakers and across several measures. We find little evidence for individual tendencies in convergence or for conversation-specific convergence tendencies across measures. However, there is consistency in convergence elicited by particular interlocutors, across measures and across conversations. These results indicate that the context of the particular conversational partner plays a larger role than speaker in influencing degree of convergence and suggest that it will be more fruitful to look for individuals who elicit convergence rather than individuals who exhibit convergence.

Keywords: convergence, consistency, individual differences, social factors, multiple measures
1 Introduction

Convergence is the increase in similarity between interlocutors; it has been demonstrated in many measures, both linguistic and non-linguistic. Variation in degree of convergence has been found in many studies, with both social and personal predictors, but little work has tested whether differences are consistent for an individual or a pair of interlocutors across measures or across tasks. Most studies include each individual in a single interaction and examine only a single predictor, so it is unclear whether variation is more due to personal factors of individuals or situational factors of the particular interaction. In this paper, we present a set of studies probing individual convergent tendencies, individual tendencies in eliciting convergence, and conversation-specific tendencies, comparing across measures. We propose that degree of convergence is mediated more by characteristics of the situation than characteristics of the speaker. Consistency of convergence elicited by the same interlocutor across measures in our data supports a unified mechanism driving convergence in different measures, but the lack of consistency across measures within a conversation indicates that there must be different mediating factors, if there is a consistent underlying mechanism of convergence.

Degree of convergence seems to be influenced by personal characteristics of the participants, such as social investment (Natale, 1975); Yu et al. (2013) regard it as evidence for individual differences in language processing, based on the relationship between Autism Quotient scores and convergence. If there are broad social or cognitive differences driving differences, their effects should be consistent across different measures of convergence. However, there is limited work testing consistency in individual convergent tendencies across tasks or in conversations with different partners, and even less work comparing across measures.

Degree of convergence is predicted by aspects of the interaction and the relationship between the interlocutors (e.g. Gregory and Webster 1996; Pardo et al. 2012; Bane et al. 2010), but most studies examining such effects do not compare convergence across measures within conversations. If variation is driven by situational or interactional factors, there should be consistency across similar conversations and between participants within the same conversation, as long as the situational effects are symmetrical. Moreover, if convergence in different characteristics reflects the same un-
derlying process, situational influences should be correlated across different measures within a conversation, unless there are mediating factors such as amount of exposure that drive differences across measures. One particular situational factor that might shape conversational interactions is the interlocutor. Some individuals might elicit more or less convergence, either based on their behavior, or based on what the listeners know or think they know about them; previous work has demonstrated some such effects (Babel, 2009; Yu et al., 2011; Pardo et al., 2017).

1.1 How Convergence is Measured

Convergence can be measured with two basic types of comparisons. It can be observed in overall decrease in distance between interlocutors or between an individual and a model talker, comparing productions from before exposure to that speaker and during or shortly after exposure, from shadowing tasks or conversation (e.g. Goldinger 1998; Pardo et al. 2017; Babel 2009). Convergence can also be observed in speakers’ synchrony over time within shared conversations (e.g. Schweitzer and Lewandowski 2013; Levitan and Hirschberg 2011).

Some patterns of synchronous change among interlocutors may result from external influences that cause the same shift in both interlocutors; some work includes comparisons with other speakers performing the same task, to control for task-related effects (e.g. Sanker 2015; Oben and Brône 2016). However, these comparisons do not control for effects that the particular conversation might have on both interlocutors, such as excitement about the conversational topic. Such effects seem to be responsible for the significantly greater similarity of an interlocutor to a speaker as compared within a conversation than as compared to that speaker’s characteristics from other conversations (Gregory and Webster, 1996).

Nonetheless, convergence is apparent even when comparing speakers to their interlocutors’ productions in other conversations (Cohen Priva et al., 2017). Comparing speakers to their interlocutors’ baselines provides a way to examine the extent to which each partner is converging, while comparing change in distance between interlocutors before and after their interaction or from the beginning to the end of their interaction does not measure how much each partner has shifted.
Controlling for change made by each interlocutor, as in Cohen Priva et al. 2017, can make convergence results for conversations more parallel to convergence results from shadowing tasks, in which there is only one speaker whose productions can change. Measuring change in distance obscures changes in which speakers shift in the direction of interlocutors, but shift so far in that direction that they overshoot the interlocutor’s original value and end up with a greater final distance than their starting distance; such changes would be treated as the same as divergence. Using individuals’ baselines as predictors makes it possible to distinguish between divergence and overshooting.

1.2 Variation by Measure

Convergence is often treated as a unified phenomenon, implicitly suggesting that it will behave similarly across measures. Most analyses do not specifically make predictions for patterns across measures, though some theoretical accounts do explicitly predict a correlation of convergence across measures, at least locally, based on alignment at one level facilitating alignment at other levels (Pickering and Garrod 2004:174-175). However, there are differences in convergence depending on the characteristic used to measure it (e.g. Pardo et al. 2017; Levitan and Hirschberg 2011; Babel 2012; Sanker 2015), which may suggest that convergence is mediated by differences in exposure or by different representations for different measures. One possible explanation is cumulative priming (cf. Kaschak et al. 2011; Oben and Brône 2016); different constructions and forms have been presented a different number of times and thus the ones which were presented more have been primed more. However, these studies have not specifically examined predictions made across measures. Studies are often based on a small number of measures, without a systematic comparison of measures, so the relative behavior of different characteristics cannot be used to inform analysis of the underlying mechanism.

Convergence has been observed in a range of diverse characteristics: nonlinguistic characteristics like fidgeting (Chartrand and Bargh, 1999) and posture (Dijksterhuis and Bargh, 2001) and linguistic characteristics, including vowel formants (e.g. Babel 2009; Pardo et al. 2012), vowel duration (e.g. Sanker 2015; Pardo et al. 2017), voice onset time (e.g. Yu et al. 2013; Bane et al. 2010), F0 mean (e.g. Babel and Bulatov 2011; Pardo et al. 2017), F0 range (e.g. Vaughan 2011; Michalsky and Schoormann
lexical choice (e.g. Brennan and Clark 1996; Branigan et al. 2011), syntactic constructions (e.g. Branigan et al. 2007; Bock 1986), and speech rate (e.g. Staum Casasanto et al. 2010; Cohen Priva et al. 2017), among others. Convergence has also been measured holistically, using AXB tasks in which listeners make decisions about similarity (e.g. Goldinger 1998; Pardo et al. 2012).

A complication in comparing across measures of convergences is that there are tendencies in the types of studies that use different measures. Many convergence studies on phonologically contrastive characteristics such as vowel formants and VOT use shadowing tasks, with the same items in training and in testing (e.g. Namy et al. 2002; Babel 2009) and less frequently test these characteristics in natural conversations (though cf. Bane et al. 2010; Sanker 2015). On the other hand, non-contrastive linguistic characteristics such as lexical choice and syntactic constructions are often tested in priming paradigms using description of parallel scenes or objects (e.g. Bock 1986; Branigan et al. 2007) or conversations (e.g. Brennan and Clark 1996; Oben and Brône 2016; Reitter et al. 2006), but not explicit shadowing.

1.3 Individual Factors

Within studies, there is often variation in the degree of convergence exhibited by each participant or pair of interlocutors. This variation has sometimes been attributed to individual differences, based on the existence of personal factors that are significant predictors of degree of convergent behavior, such as higher social desirability scores (Natale, 1975), higher openness and attentional focus scores (Yu et al., 2013), smaller social networks (Lev-Ari, 2018), and tendency to compromise (Weatherholtz et al., 2014). However, many of these observed effects are small, and without retesting of individuals to establish their consistency, interpretations are at risk of a fundamental attribution error; many of these studies implicitly or explicitly assume that variation in convergence reflects inherent characteristics of the individuals. However, evidence for individual tendencies in convergence is somewhat limited, and variation may be more driven by external factors.

Some work has demonstrated individual differences in linguistic behaviors that are consistent within a limited domain, for example, categorical perception (Kong and Edwards, 2016) or lexical bias.
(Ishida et al., 2016). Such differences are sometimes interpreted as reflecting broader cognitive differences that might explain other behaviors and phenomena, such as the propagation of language change (e.g. Yu 2013; Lev-Ari 2018). Individual speakers are highly variable in their own productions; differences in individual variability has also been suggested as a possible individual difference that could make some individual more innovative (Sonderegger et al., 2017), though the high level of variability within individuals also creates noise for evaluating individual differences in convergence.

Only a few studies have tested individual consistency in convergence, with the same individuals participating in multiple conversations or experimental tasks. Within a given measure of convergence, individuals do exhibit some consistent tendencies in the same or very similar tasks, particularly with the same interlocutor or model talker (Sanker, 2015; Tamminga et al., 2018a), but these tendencies are much weaker across more dissimilar tasks such as shadowing and conversation (Pardo et al., 2018).

Accounts of individual differences would predict that different measures of convergence should exhibit parallel behavior across individuals (Weise and Levitan, 2018; Cohen Priva and Sanker, 2018); if variation in convergence results from broad differences in language processing or social interaction, an individual pattern of convergence in one measure should be predictive of that speaker’s convergence in other measures. Some measures of convergence are affected similarly by the same aspects of individuals (Natale, 1975; Yu et al., 2013; Lev-Ari, 2018; Weatherholtz et al., 2014). However, there is little work that compares convergence by each individual across measures; existing work with such comparisons has not found reliable correlations across measures. There was no tendency for individual consistency in convergence across different measures in most studies that looked for such patterns (Bilous and Krauss, 1988; Pardo et al., 2012; Sanker, 2015; Weise and Levitan, 2018), though the null result could potentially be attributed to the rather small number of participants in these studies. In studies with multiple comparisons, some correlations have appeared to be marginally significant (Rahimi et al., 2017; Sanker, 2015), but with low p-values and no clear motivation for the particular measures that correlates, which suggests that these are cases of false positives resulting from repeated measures.

The lack of clear evidence for consistency in convergence across measures and across tasks sug-
gests that weak individual tendencies in convergence, which might be observable within a given measure in a given task, are at an extremely low level, rather than a high level that would translate to similar results in different tasks and linguistic characteristics. Moreover, stronger effects such as the task and interlocutor are likely to outweigh these weak individual tendencies.

1.4 Situational Factors

External situational factors such as social characteristics and the relationship between the participant and interlocutor or model talker are significant predictors of convergence. Factors that interact with convergence include race (e.g. Babel 2009), gender (e.g. Bilous and Krauss 1988; Pardo 2006; Namy et al. 2002), dialect (Giles et al., 1991; Drager et al., 2010; Babel, 2010), perceived standardness of the model talker’s dialect (Weatherholtz et al., 2014), interlocutor status (Gregory and Webster, 1996; Bane et al., 2010), speaker’s role in the task (Pardo, 2006), positiveness of the interlocutors’ relationship (e.g. Pardo et al. 2012; Sanker 2015), and attitude towards a model talker (e.g. Babel 2010; Yu et al. 2011). The effects of social factors on convergence are not always consistent, and can interact with each other and other factors, e.g. Pardo (2006) found more converge exhibited by men, but only for the speakers whose role in the task was to receive instructions, and Namy et al. (2002) found more convergence among women, but only for one model talker.

Many of these social factors capture real or perceived characteristics of the interlocutor, though others are about the particular interaction between the two speakers. Even within shadowing tasks without explicit information provided about the model talker, different talkers can elicit different degrees of convergence (Pardo et al., 2017), though this effect might be a combination of effects of each speaker and of each recording. Convergence can also depend on the starting distance between speakers within a measure; individuals with more similar productions have less room for convergence and will thus often exhibit less convergence than individuals with more distinct starting points (Kim 2012:80; Babel 2010:453), though greater distance can also be inhibitory across dialects or native languages (Kim et al., 2011).

Many of the situational effects of conversations are symmetrical between interlocutors; speak-
ers tend to converge more when their interlocutors converge more, though sometimes convergence is driven more by one interlocutor than the other. Similarity in convergent behavior of both partners may reflect the interlocutors’ relationship, as the positiveness of the interaction affects both partners, and positive interactions produce more convergence. A lack of convergence from one partner can result in less convergence from the other partner, as Gregory et al. (1997) demonstrate with the lack of convergence exhibited by speakers whose partners are receiving filtered transmissions of their speech; individuals receiving high-pass filtered input do not hear F0 and thus cannot converge to F0, but their partners, who are receiving unfiltered input, also do not converge. The relationship between convergent behavior of partners is also apparent in the change of each individual within unfiltered conversations, though the results are not consistent across studies; Sanker (2015) found that a speaker’s degree of change in a measure was predictive of her partner’s degree of change in that measure, though Pardo et al. (2018) found no correlation.

Convergence across measures within a conversation or shadowing task is less clearly related. In the same task, there can be differences in overall convergence in different measures (e.g. Babel 2009; Sanker 2015) and also differences in effects of conditioning factors (e.g. Bilous and Krauss 1988; Pardo et al. 2017). However, little work has systematically tested whether the convergence exhibited by each speaker or interlocutor pair in one measure was predictive of their convergence in other measures within the same task. In a comparison across a small number of conversations, Sanker (2015) found only a very weak trend towards positive correlations across measures. With a larger dataset, Weise and Levitan (2018) found no correlation in convergence across measures. Cohen Priva and Sanker (2018) found a correlation only between closely related measures: F0 median and F0 variability.

1.5 This Study

Using data from the Switchboard corpus (Godfrey and Holliman, 1997), we examine convergence in 6 different measures, some of them related and others unrelated. We present comparisons of consistency and convergence across measures, individual tendencies for convergence within and across measures, per-conversation tendencies for convergence, and tendencies of an individual to elicit con-
All the models presented here predict speakers’ performance in a conversation using their own baselines (measured as their performance in other conversations) to model how consistent they were, and their interlocutors’ baselines (their performance in other conversations) to model how convergent they were, building on the approach taken by Cohen Priva et al. (2017). Other predictors were added to test factors influencing raw productions and influencing degree of convergence. In this study, we will examine only conversation-level convergence effects, and not time-aligned effects.

Individual consistency within a characteristic is likely to depend on the extent to which each measure reflects personal habits, phonologically dictated targets, or physical constraints. F0 measures should exhibit consistency reflecting physical limitations of each speaker, while characteristics that are strongly influenced by topic and type of discourse (lexical choice, sentential conjunction) should exhibit the lowest stability.

Based on treatment of convergence as a unified phenomenon, we might expect that all measures should exhibit similar degrees of convergence. However, previous work has demonstrated that this is not the case, so the more relevant question is which characteristics will exhibit the most convergence, to the extent that they are measured in comparable ways. Given that convergence is measured here relative to interlocutors’ baselines, rather than interlocutors’ productions within the shared conversation, convergence should be strongest in the measures which are most dictated by individual tendencies in usage rather than effects of environment. Cumulative priming predicts that convergence should be strongest in the measures for which speakers received the most exposure, for example, high convergence for F0, which is present nearly continuously, and less convergence for use of sentential conjunctions, because there are fewer tokens of sentence beginnings.

Variation in convergence across participants has been observed previously. If the variation is due to broad individual differences in perception or social interaction (cf. Yu et al. 2013; Natale 1975), with one underlying mechanism that drives convergence across characteristics, tendencies in convergence should be apparent across different conversations which each participant had, and inclusion of these per-speaker tendencies for convergence should significantly improve models for production of each
characteristic. If there is no unified phenomenon of convergence across measures (cf. Weise and Levitan 2018) and variation in convergence is due to other factors, or is low-level rather than broad, there should be no effect of including individual tendencies for convergence as a factor, either in a model for a single linguistic characteristic or in a model encompassing multiple characteristics.

However, in many studies, each individual only participated in a single conversation or listening task, so variation across participants could be due to differences between interactions rather than differences between individuals. Influences of various situational factors in convergence have previously been demonstrated (e.g. Gregory and Webster 1996; Drager et al. 2010; Pardo et al. 2012). In Study 2, we also look for per-conversation variation in degree of convergence, that is, tendencies of particular conversations to be more or less convergent, across both speakers and across measures. If convergence is mediated by situational differences at the conversation level, there should non-negligible variation in this property. Otherwise, if variation in convergence is due to other factors, there should be no effect of including conversation-specific tendencies for convergence as a factor.

Situational factors may be driven primarily by characteristics of the interlocutor. Influences of particular characteristics of the interlocutor on degree of convergence have been previously demonstrated (e.g. Yu et al. 2011; Babel 2009; Weatherholtz et al. 2014), as well as less-conditioned variation based on the speaker or the recordings of each speaker (Pardo et al., 2017). In Study 2 we also look for per-interlocutor variation in degree of convergence, that is, tendencies of particular interlocutors to elicit convergence, across measures. If variation in convergence is due to situational differences driven by the interlocutor, there should non-negligible variation in this property. If interlocutor is the only situational factor influencing degree of convergence, inclusion of interlocutor as a factor should eliminate any per-conversation effects.

2 Methods Overview

2.1 Corpus

The data for this study is the Switchboard Corpus (Godfrey and Holliman, 1997), a large collection of telephone conversations. Each speaker was randomly paired with other speakers and given a topic
for each conversation, providing a large corpus of natural speech data for many speakers in similar conversations with several different partners; recordings include associated data providing speaker identification that can be used to compare the conversation with other instances of that speaker. Each side of the conversation is a distinct recording, so measurements can reliably be taken for each speaker separately.

Each conversation has associated information quantifying the clarity of the recording. To ensure reliable acoustic measurements (F0 median and variance), calls were omitted if they had high levels of background noise, echoing, or other issues, as indicated in the annotations of these calls. This resulted in a total of 464 speakers in the data that was used for acoustic measures. For measurements that did not depend on acoustic form, no conversations were omitted, so these measures were based on 518 speakers. Conversations took 6:20 minutes on average (the median was 5 minutes). Word boundaries were based on the manually corrected word annotations produced at MS State (Harkins et al., 2003). The word annotations allow measurement of word duration.

### 2.2 Measuring Convergence

In shadowing tasks (e.g. Goldinger, 1998; Babel, 2012; Pardo et al., 2017), speakers (S) are exposed to a pre-recorded “interlocutor” (I) whose speech they need to repeat, with or without a request to sound like them. Speakers are measured before the exposure (Sb) and after the exposure takes place (SI). In broadest terms, convergence would be any change in the difference between Sb and SI that makes SI more I-like. If approximated as a linear combination as in (1), convergence could be measured as the value of βI, the relative importance of the interlocutor in predicting the performance of speakers after the interaction, relative to their consistency (self-correlation), which would be measured in βSb. This model can be described as follows: Speakers’ performance results from a combination of their self-consistency, the performance of their interlocutor, and noise.1

\[
S_I \approx \beta_0 + \beta_{S_b} S_b + \beta_I I + \epsilon
\]

1Schweitzer and Lewandowski (2013) use a coefficient for the interlocutor alone, and model \( \beta_{S_b} S_b \) as a random intercept.
One limitation of this approach is in measuring individual differences in convergence. In many experimental designs, only a single set of before and after values are available for each speaker, which is not enough to estimate all the parameters in the equation separately for every speaker. Rather than this sort of model, many studies reduce the dimensionality of the problem by modeling convergence as an increase in the similarity between $S_I$ and $I$ relative to the similarity between $S_b$ and $I$. One approach to measure similarity in this way is subjectively, as in AXB designs (e.g. Goldinger, 1998; Pardo et al., 2012), which yield holistic (not measure-specific) binary judgements. Another common approach is to model similarity as the absolute difference between the speaker and interlocutor performance (e.g. Babel, 2012; Pardo et al., 2013), in which convergence is modeled as in (2), a difference in (absolute) differences (DID).

\[
\text{(2)}

DID := |S_I - I| - |S_b - I|
\]

The DID approach has several disadvantages. It does not distinguish between differences that follow from lack of convergence, and differences that follow from over-convergence. Thus, if $S_b = 5$ and $I = 4$, then $S_I = 5$ and $S_I = 3$ would be equally non-convergent (DID=0), despite the former exhibiting a true lack of movement, and the latter exhibiting over-convergence. Furthermore, the elimination of terms for speakers’ consistency ($\beta_{S_b}$ above) and the error term means that DID estimates would be less reliable when the difference between the speaker and the interlocutor is initially small. In such cases, lack of self-correlation (or noisy data) would be more likely to be interpreted as divergent than for speakers whose initial performance is further from the interlocutor. In addition, measuring speakers’ baselines using a small number of words, particularly if they are produced in unnatural conditions such as reading word lists, can create noise due to not accurately capturing speakers’ typical speech. Section 1 in the supplementary materials uses one of the models in Study 1 to demonstrate how the DID approach could lead to superfluous correlations between initial values and convergent behavior.

Some studies use acoustic manipulations to target convergence in particular characteristics by creating extreme values in the model talker (e.g. lengthened VOT in Nielsen, 2011; Yu et al., 2013).
These manipulations ensure a relatively large starting distance between the speaker and the model talker, which makes over-convergence unlikely and reduces the likelihood that random variability within speakers will exceed the size of the convergence effects.

Some studies compare two conditions, rather than using DID, and test whether productions in the two conditions differ in the expected direction. This sort of analysis is typical in syntactic and lexical priming tasks, which often only present the target words or constructions, with the synonymous competitor never appearing (e.g. Branigan et al., 2011; Bock, 1986; Brennan and Clark, 1996).

Measuring convergence in conversations introduces additional complications. Crucially, there is no longer direct access to the interlocutor’s baseline. Both the speaker and the interlocutor may converge, so comparing $S_I$ to the interlocutor’s actual performance could yield false positives, even in the absence of actual convergence. For example, if one speaker always shouts and never listens (and therefore cannot converge), an interlocutor who shouts back would make the speaker appear convergent, and not capture the unequal contributions of each speaker to that convergence. Similarly, if the subject matter requires a particular speaking style (e.g. using low frequency words or slower speech rate), both parties would seem to converge if they adopt that speaking style, even if they ignore each other’s performance. A simple solution to this issue is to obtain a baseline for the interlocutor too. In the Switchboard corpus, in which speakers interact with multiple different interlocutors, a natural way to approximate speakers’ and their interlocutor’s baselines is to average their performance from other contexts, in which they were interacting with other conversation partners, as in Cohen Priva et al. (2017).

While the issues with measuring convergence as change in distance are likely to average out across participants and not obscure the overall result, the degree of noise may be enough to obscure the results for each participant and hinder analysis of individual differences. Our method makes it possible to measure each individual’s contribution to convergence, as well as providing a strong baseline and establishing each speaker’s self-consistency, which reduces some of the risks of measurement artifacts.

First, each measure in each individual conversation is processed separately. For every measure, multiple individual values (e.g. F0) in each conversation side are summarized to a single statistic (the
median, inter-quartile range, or average, depending on the measure). Thus each conversation side is represented by a single value per measure. This step is illustrated for speech rate values in the top three and bottom three boxes in Figure 1.

Second, for each measure, the single values of all the speaker’s conversations are summarized to a single data point, which contains for each speaker and interlocutor, (a) the speaker’s performance during the conversation (equivalent to $S_I$ above), (b) the speaker’s mean performance in other conversations, to be used as the speaker’s baseline ($S_b$), and (c) the interlocutor’s mean performance in other conversations, to be used as the convergence target $I$.

Finally, the multiple data points per measure are collected to create the set of data points that are used to predict convergence, as illustrated for speech rate in Figure 2. While this method may decrease the amount of actual convergence captured, convergence is still apparent when tested in this way. Establishing reliable baselines depends on having a large corpus, so that baselines are averaged across enough conversations to not be thrown off by the particular characteristics of any particular conversation; the median number of conversational partners for speakers in our dataset was 9 (Q1:4, Q3:12).

2.3 Measures

We included six measures for convergence. Our goal was to have a broad range of measures. Four of the six, F0 median, F0 variance (IQR), speech rate, and *uh:um* ratio were examined in Cohen Priva and Sanker (2018). Two measures were added to both expand the range of measures used and to involve additional linguistic aspects: lexical information rate and sentential conjunction. All six are described below.

**F0 median.** F0 is the rate at which vocal folds are vibrating, which occurs during vowels and other voiced sounds; this frequency is determined by the thickness and tension of the vocal folds. Speakers’ baseline F0 is largely determined by anatomical characteristics (Titze, 1989; Krook, 1988), but there are additional factors that influence speakers’ usage of the ranges which they are physically capable of (Jessen et al., 2005; Scherer et al., 1991). Patterns in perception suggest that logarithmic or similar
Figure 1: An illustration of the first two steps in the procedure, as they apply to speech rate. First (top three and bottom three boxes), the mean of multiple speech rate values (log observed over expected) is taken, yielding a single point per conversation side. Second, for each conversation side, the speaker’s performance provides the predicted value, the mean of the speaker’s performance in other conversations is used as the speaker’s baseline ($S_b$), and the interlocutor’s performance in other conversations is averaged to yield the interlocutor’s baseline ($I$).

scales best characterize $F_0$ in speech (Stevens et al., 1937; Ladd, 1996), so all frequency measures were taken in Mels. Much convergence work on $F_0$ has used Hertz (e.g. Pardo et al., 2017; Gregory et al., 1997), though some work has used non-linear scales such as ERB and Mel (e.g. Babel and Bulatov, 2011). Following the majority of previous work, we do not scale $F_0$ within speakers, given the lack of evidence that speakers converge to their interlocutors’ relative $F_0$ rather than their absolute $F_0$. $F_0$ was measured in Praat using the default settings (To Pitch: 0.001, 75, 300); while these settings sometimes fail to capture instances of creaky voice, they are optimal for minimizing pitch tracking
errors and capturing the F0 range of a large number of speakers. F0 was measured separately for each word, which is more accurate than measuring it for entire conversations and allowed measurement of speakers’ F0 variability within conversations. Tokens beyond density minima at either end of the distribution were excluded, to minimize the amount of data resulting from pitch tracking errors. While this exclusion process might also exclude some cases of actual extreme F0, an investigation of some of the apparent outliers confirmed that most of them are the result of erroneous pitch tracking halving or doubling the actual F0.

**F0 range** was measured as the log of the ratio of the 75th percentile to 25th percentile of F0 mea-
measurements in Mels. Using the quotient rather than the difference (cf. Michalsky and Schoormann, 2017) was aimed at reducing the artificial correlation between F0 median and F0 range; using a difference or a standard deviation will make the range will scale up in proportion to the center of the distribution in a way that does not align with listener perceptions or production patterns. Using the quartiles reduced the sensitivity of the measurements to outliers and the outlier exclusion method employed in the F0 data.

**Speech rate** was remeasured using the data reported by Cohen Priva et al. (2017, Study 2), using only the factors of interest: Speakers’ baseline and interlocutors’ baseline, so that all the measures of convergence reported in this paper are based on exactly the same model structure and can be directly compared. Speech rate in a conversation was measured as the mean log speech rate of individual utterances. Point-wise speech rate was measured as the actual utterance duration (including pauses) divided by the expected utterance duration. Expected utterance duration was calculated as the sum of the predicted durations of words in the utterance, each calculated as the predicted value of a linear regression using the median duration of that word in the entire corpus, the length of the utterance, and the distance from the end of the utterance. Unlike the previous measurements, speech rate was calculated based on hand-corrected values. Raw measures of speech rate, such as number of segments or syllables per second, do not control for many factors that affect word duration, such as different inherent durations of different segments, effects of stress on syllable duration, and compression at the level of the syllable, word, and utterance (White, 2002). Such measures similarly fail to account for frequency effects (e.g. Bell et al., 2009). The calculation of point-wise speech rate explicitly treats each word as having a different baseline duration, in order to address these effects, which makes it a more accurate approximation of speech rate than measures based simply on the number of syllables or segments.

**uh:um ratio**, the first non-phonetic measure we used, was measured as the log odds of *uh* vs. *um*, a choice which has been attributed to processing factors (e.g. Clark and Fox Tree, 2002), but is also influenced by other factors such as gender (Acton, 2011). Log odds were calculated as the predicted values plus the residuals of a logistic regression between the number of *uh* uses and *um* uses in each
This method was used because it applies even when only *uh* or only *um* is used by a speaker in a particular conversation. Convergence has been observed previously in filler frequency (Bečuš et al., 2014) and acoustic characteristics of the realization of these forms (Bečuš, 2009). The existence of convergence in usage of *uh* and *um* suggests that the choice between them is influenced not just by communicative or processing factors, but also by social and stylistic factors. Individual consistency in their usage would provide further evidence for this function.

**Lexical information rate.** Though *uh* and *um* were argued to have a meaning (Clark and Fox Tree, 2002), it is not clear that their usage involves lexical retrieval (for instance, Bell et al., 2009, argue that function words do not involve lexical retrieval). To see whether lexical choices are also influenced by convergence, we chose to study the effect of convergence on lexical information rate. Another motivation to focusing on lexical information rate is that it has been shown to correlate with speech rate (fast talkers tend to use lower lexical information rate, Cohen Priva, 2017). Lexical information rate was measured following Cohen Priva (2017), as the mean negative log unigram predictability (the unigram entropy) of non-function words used in each conversation. Word counts were estimated using a combination of the Buckeye (Pitt et al., 2007), Fisher (Cieri et al., 2004, 2005), and Switchboard (Godfrey and Holliman, 1997) corpora. Word unigram probabilities were approximated by dividing each word’s count by the sum of word counts. Function words and frequent filler items were excluded. The probabilities were then negative log transformed and averaged across the entire conversation side to yield speakers’ unigram lexical information rate. Unigram lexical information rate captures the diversity and rarity of the words used by speakers. There are several ways that speakers may converge in lexical information rate. They may shift to more or less formal language registers or shift towards a shared specialized lexicon. To focus on that possible property, rather than on shifting choices in frequent words (e.g. expressions such as *you know, that’s right*), we excluded the top 1% most frequent words.2

**Sentential conjunction.** Coming up with a syntactic phenomenon that can be reliably measured in Switchboard is challenging due to the largely word-level annotation of the corpus. This limitation

---

2We verified that excluding the most frequent words does not have a substantial effect on the model results.
ruled out established convergence measures such as active:passive voice and dative:double accusative (cf. Bock, 1986; Bernolet et al., 2016; Hwang and Chun, 2018; Branigan et al., 2007). One construction that does meet our criteria for use with this corpus is the use of and to connect sentences, as in (3). In these cases, and is not strictly required, and the meaning of the sentence would change little if and were omitted, so sentences with and without this initial conjunction can be compared. These sentence-initial conjunctions have been described as discourse markers; they make clear the continuity between a statement and the preceding discourse. Their presence or absence is partially predicted by the type of discourse, with more such connectives in narratives (Schiffrin, 1986; Dorgeloh, 2004). Despite the slightly different pragmatic distribution of sentences with and without these conjunctions, they are basically equivalent, like other constructions that are used in syntactic priming studies. For each conversation side, we counted the number of times “and PRON-NOM” appeared, where PRON-NOM could be any of I, she, he, or they (~22,360 tokens). We excluded cases in which the preceding word was a given name (~15 tokens), as in (4), sequences containing you and I (~140 tokens), and word sequences of “my kinship-term and I” (~400 tokens), to exclude cases such as (5). Such constructions comprise the majority of “and PRON-NOM” sequences that are not cases of sentence conjunction. As with uh:um ratio, we used a logistic regression to calculate the log odds between PRON-NOM with and without a preceding and, providing a comparison of pronominal sentences with or without sentential conjunction. There were ~242,700 uses of PRON-NOM, of which ~21,800 were preceded by and.

(3) ...I certainly wouldn’t object to it and I think random [testing for drugs] is probably, you know, the only really fair way ...(SW2638A)

(4) ...that Dan and I are going to ...(SW3323B)

(5) ...in the old days when my wife and I both worked ...(SW4238A)
3 Study 1: Stability and Convergence

3.1 Introduction

The goal of this study is to establish whether speakers converge along several linguistic aspects, meant to span from phonetic properties to syntactic properties. Four of these measures replicate existing findings (Cohen Priva and Sanker, 2018), and two were added to examine a more diverse set of linguistic features. Other than re-establishing that the corpus is sufficient for detecting convergence along the six properties, we were interested in studying the degree in which linguistic properties differ with respect to speakers' consistency and likelihood to converge, and whether those differ for different linguistic properties. Crucially, we wanted to look for evidence of individual variance in likelihood of converging.

3.2 Methods and Materials

3.2.1 Statistical Models

For each individual measure, we built a mixed effects linear model in R (R Core Team, 2018) along the lines of (1) as illustrated in Table 1, predicting speakers' performance using their own baseline and their interlocutors' baseline as the two fixed predictors. Speakers' consistency is captured in the coefficient of their baseline. Speakers' convergence is captured in the coefficient of the interlocutor's baseline. The coefficients were modeled using the lmerTest package (Kuznetsova et al., 2015), which encapsulates lme4 (Bates et al., 2015), but calculates degrees of freedom (Satterthwaite approximation) and uses them to provide p-values.

A random intercept per speaker was included for completeness, even though the fixed effects for the speakers’ baseline is supposed to capture the same variance. A random intercept per interlocutor was also added, despite a similar limitation. The advantage of a random intercept per interlocutor is in handling per-interlocutor trends that may not fall under convergence per se, such as speaking more slowly with interlocutors who demonstrate difficulties understanding fast speech, which does not necessarily align with the speech rate that those individuals use themselves. Additionally, we included
random intercepts for the conversation and the topic, as listed in the metadata of the Switchboard corpus. The former is meant to capture the variance that may be associated with a particular conversation, independent of the tendencies that the speakers involved might exhibit in other conversations, and the latter is meant to capture the variance associated with topic-specific biases (e.g. perhaps faster speech when discussing sports). Finally, we included per speaker and per interlocutor random slopes for the interlocutors' baselines. These are our variables of interest: If some speakers are more likely than others to use the interlocutor's performance to determine their own performance (“natural followers”), we would find per speaker variance in the slope of the interlocutor’s baseline. Similarly, if some interlocutors are more likely to elicit convergence (“natural leaders”), we would find variance by-interlocutor in the slope of the interlocutor’s baseline. Since the random effects structure is complex and overspecified (we do not expect the speakers’ intercept to capture variance), we fit each random effect separately, forcing lme4 not to compute the covariance between speaker intercept and slopes, and between interlocutor intercept and slope (cf. Bates et al., 2015). To clarify our modeling choices, we provide the lme4 formula in Table 1, as its syntax has become the de-facto standard for model fitting.

All dependent variables and predictors were standardized, to allow for an easier comparison across different models. This means that if the absolute value of a coefficient is very close to 1, it is highly predictive of the actual performance. Coefficients that are close to zero would indicate that (a) other factors explain the observed variance better or (b) the observed values are noisy, either inherently or due to limitations in the measurement. For the fixed effects, p-values are provided by the lmerTest package as part of the final model. For the random effects, p-values are based on model comparison, contrasting a model that uses a variable with a minimally different model that does not use that variable.

3.2.2 Removal of Per-speaker Intercepts

In all models except the one for sentential conjunctions, per-speaker intercepts were assigned zero variance. In the sentential conjunction model, the values assigned to each speaker were highly correlated with the speaker's baseline (Pearson $r = .974$); the speaker baseline coefficient behaved as a
### Table 1

<table>
<thead>
<tr>
<th>lme4 syntax</th>
<th>explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>speaker ~</td>
<td>Speaker's performance in a conversation (S_I)</td>
</tr>
<tr>
<td>1</td>
<td>Intercept, expected to be zero (\beta_0)</td>
</tr>
<tr>
<td>+ speaker.baseline</td>
<td>Speakers' baseline: captures consistency (\beta_{bs} S_b)</td>
</tr>
<tr>
<td>+ interlocutor.baseline</td>
<td>Interlocutors' baseline: captures convergence (\beta_I I)</td>
</tr>
<tr>
<td>+ (1</td>
<td>speaker)</td>
</tr>
<tr>
<td>+ (1</td>
<td>interlocutor)</td>
</tr>
<tr>
<td>+ (1</td>
<td>conversation)</td>
</tr>
<tr>
<td>+ (1</td>
<td>topic)</td>
</tr>
<tr>
<td>+ (0 + interlocutor.baseline</td>
<td>speaker)</td>
</tr>
<tr>
<td>+ (0 + interlocutor.baseline</td>
<td>interlocutor)</td>
</tr>
</tbody>
</table>

supresor for the random intercept, and was assigned a negative coefficient, despite being positively correlated with the predicted variable (Pearson \(r=.4\)). These unreliable outcomes are to be expected, given that speakers are highly consistent in their performance and their consistency is modeled directly using their individual baselines (and which, unlike the random intercept, are not required to be normally distributed). We therefore refitted the models without the per-speaker intercept. The results of the redundant models are provided in Section 2 of the supplementary materials.

### 3.3 Results

In every model, speakers were highly consistent, as shown in Table 2, though consistency differed across the characteristics. The high value associated with F0 median likely follows from pitch being highly dependent on physiological properties and variable across individuals. The unusually low consistency for sentential conjunction rate likely emerges from language structure being more predictive of word choice than each speaker's tendencies. Figure 3 visualizes the correlation between speakers' baseline and speakers' performance. We normalized by measure, to facilitate comparisons across measures, but did not normalize by speaker, so the distributions within measures are unaltered. Thus, F0 preserves a strongly bimodal distribution consistent with the sex-based differences. Most convergence studies using F0 similarly measure convergence in absolute values (e.g. Pardo et al. (2017); Babel
and Bulatov (2011)), though some work has normalized by gender (Weise and Levitan, 2018); we follow the former works, given the lack of clear evidence for perceptual normalization shaping convergence.

Table 2: Speaker consistency coefficients in Study 1, per measure

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 median</td>
<td>0.971</td>
<td>0.0039</td>
<td>3434</td>
<td>248</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>F0 variance</td>
<td>0.675</td>
<td>0.0119</td>
<td>3622</td>
<td>57</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Speech rate</td>
<td>0.800</td>
<td>0.0086</td>
<td>4336</td>
<td>93</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>uh:um ratio</td>
<td>0.787</td>
<td>0.0090</td>
<td>4501</td>
<td>87</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Lexical information rate</td>
<td>0.645</td>
<td>0.0095</td>
<td>4490</td>
<td>68</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Sentential conjunction</td>
<td>0.391</td>
<td>0.0131</td>
<td>4677</td>
<td>30</td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>

Convergence, as indicated by the degree to which the interlocutor’s baseline predicted for the variable, accounted for far less variance than the speaker’s baseline, as Table 3 shows. However, significant convergence was found in every model. Convergence predicted the least variance for F0 median, which is likely the result of speakers’ extremely high consistency in this characteristic. Figure 4 visualizes the correlation between interlocutors’ baseline and speakers’ performance, after residualizing the effect of the speakers’ own baseline.

Table 3: Convergence coefficients in Study 1, per measure

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 median</td>
<td>0.0179</td>
<td>0.0049</td>
<td>181</td>
<td>3.7</td>
<td>0.00033</td>
</tr>
<tr>
<td>F0 variance</td>
<td>0.0929</td>
<td>0.0140</td>
<td>143</td>
<td>6.6</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Speech rate</td>
<td>0.0477</td>
<td>0.0097</td>
<td>219</td>
<td>4.9</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>uh:um ratio</td>
<td>0.0320</td>
<td>0.0110</td>
<td>147</td>
<td>2.9</td>
<td>0.00433</td>
</tr>
<tr>
<td>Lexical information rate</td>
<td>0.0612</td>
<td>0.0110</td>
<td>224</td>
<td>5.5</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Sentential conjunction</td>
<td>0.0405</td>
<td>0.0148</td>
<td>152</td>
<td>2.7</td>
<td>0.00699</td>
</tr>
</tbody>
</table>

Table 4 provides the standard deviation and model comparison-based p-values for the three random intercepts that remained in the model. Conversation significantly contributed to predicting the speaker’s performance in every model. Because the conversation-level intercept applies to both conversation sides, it captures elements in the conversation in which the two participants vary together. This is not convergence in the strict sense, as it may be driven by external characteristics and not
Figure 3: Speakers’ consistency, in terms of the relationship between their baseline and actual performance, by measure. The line shows the correlation and the grey areas (hardly visible) highlight the confidence interval. Every conversation side is a single data point. Values were normalized within each measure.
Figure 4: Speakers’ convergence, in terms of the relationship between their interlocutors’ baseline and actual performance, residualized to remove the effect of their own baseline, by measure. The line shows the correlation, and the grey areas highlight the confidence interval. Every conversation side is a single data point. Values were normalized within each measure.
just speakers’ influence on each other. Similarly, topic identity contributed to every model, indicating that different topics have typical values for all measures (e.g. higher pitch, more uh values). The interlocutor had a significant effect on speakers’ performance only in the speech rate model, though marginal effects were found for several of the other characteristics. This suggests that the main influence that interlocutors have on speakers’ performance is via convergence, as discussed below, rather than absolute effects.

Table 4: Random intercept SD and model-comparison p values in Study 1

<table>
<thead>
<tr>
<th></th>
<th>Interlocutor</th>
<th>Topic</th>
<th>Conversation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SD</td>
<td>p</td>
<td>SD</td>
</tr>
<tr>
<td>F0 median</td>
<td>0.036285</td>
<td>0.0680</td>
<td>0.0218</td>
</tr>
<tr>
<td>F0 variance</td>
<td>0.099476</td>
<td>0.0530</td>
<td>0.1227</td>
</tr>
<tr>
<td>Speech rate</td>
<td>0.090704</td>
<td>0.0012</td>
<td>0.1288</td>
</tr>
<tr>
<td>uh:um ratio</td>
<td>0.069665</td>
<td>0.1384</td>
<td>0.0509</td>
</tr>
<tr>
<td>Lexical information rate</td>
<td>0.000538</td>
<td>1.0000</td>
<td>0.4012</td>
</tr>
<tr>
<td>Sentential conjunction</td>
<td>0.069681</td>
<td>0.4568</td>
<td>0.2145</td>
</tr>
</tbody>
</table>

Table 5 provides the standard deviation and model comparison-based p-values for the per-caller and per-interlocutor random slopes for the interlocutors’ baselines in the six models. These values indicate individual speakers’ tendencies to converge and to elicit convergence, respectively. Per-caller random slope was only significant for uh:um convergence, and per-interlocutor random slope was only significant for lexical information rate. These results suggest that the individual differences in speaker and interlocutor are too weak to be detected by the method used here. Recent work has similarly generally found a lack of consistency in the degree of convergence exhibited by individual speakers across tasks (Pardo et al., 2018), though they do exhibit some consistency within a measure in the same task or very similar tasks (e.g. Sanker, 2015; Tamminga et al., 2018b). Pardo et al. (2017) found that some model talkers consistently elicited more convergence than others in a shadowing task, though crucially this study involved the same particular recordings of these individuals, and the same task for the listeners.
Table 5: Random slope SD and model-comparison $p$ values in Study 1

<table>
<thead>
<tr>
<th></th>
<th>Per-caller convergence slope</th>
<th>Per-interlocutor conversation slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SD</td>
<td>$p$</td>
</tr>
<tr>
<td>F0 median</td>
<td>0.0237</td>
<td>0.265</td>
</tr>
<tr>
<td>F0 variance</td>
<td>0.0546</td>
<td>0.447</td>
</tr>
<tr>
<td>Speech rate</td>
<td>0.0214</td>
<td>0.825</td>
</tr>
<tr>
<td>uh:um ratio</td>
<td>0.0734</td>
<td>0.017</td>
</tr>
<tr>
<td>Lexical information rate</td>
<td>0.0000</td>
<td>1.000</td>
</tr>
<tr>
<td>Sentential conjunction</td>
<td>0.0637</td>
<td>0.326</td>
</tr>
</tbody>
</table>

3.4 Discussion

The choice of measure used to test convergence can have a large impact on results. In our data, convergence was most apparent within F0 range, though this measure does not consistently stand out as the strongest measure of convergence across different studies (cf. Vaughan, 2011; Michalsky and Schoormann, 2017). Convergence was very strong in all of the measures. Differences in convergence by measure have also been observed elsewhere (Pardo et al., 2017; Sanker, 2015; Bilous and Krauss, 1988), though it is not clear whether or not there are consistent patterns of which measures exhibit the strongest convergence or how those patterns are influenced by the task. Different amounts of exposure to each characteristic within the same speech may be in part responsible for the variation.

There are differences in the size of the coefficient for speaker baseline as a predictor for different measures. However, the measure of individual consistency can result from several factors: consistency of each individual, distinctness of each individual, and amount of data for each measure. While large coefficients for speaker baseline are unambiguous, because they require individual consistency in addition to distinctness of individuals, lower consistency can result from multiple factors, so the different measures are not necessarily directly comparable. F0 median has an extremely high coefficient, which is consistent with the physical differences driving it. Smaller coefficients for lexical frequency and sentential conjunctions are likely due to the strong influence of conversational topic on word choice and the influence of discourse structure on use of sentence-linking conjunctions, making each
speaker less consistent and driving the same patterns across speakers.

Some work has suggested that characteristics with more variable realizations are more likely to exhibit convergence (e.g. Babel, 2009, pp. 141–142), based on suggestions that convergence develops out of naturally occurring variation in speech (e.g. Pierrehumbert, 2002). However, within our results, there was no such pattern; measures with greater individual consistency did not exhibit less convergence. Most notably, individual consistency was very high for F0 median, but it still exhibited significant convergence.

4 Study 2: Convergence by speaker, by conversation, and by interlocutor

4.1 Introduction and materials

Individual differences in convergence proved difficult to find. Per-speaker and per-interlocutor slopes for convergence were assigned little or no variance when modeled using mixed effects models. Cohen Priva and Sanker (2018) found no correlation between individual differences for speakers across four measures; when we extended this approach to the 6 individual measures in the current set of studies, we only found weak correlations in individual differences between closely-related measures (F0 variance and median). Given the limited power of such pairwise comparisons, they are not reported here.

Previous studies have found individual differences in various perceptual tasks (e.g. Johnson et al. (1987); Kong and Edwards (2016); Ishida et al. (2016)), and some work has found personal characteristics such as Autism Quotient (Yu et al., 2013) and social desirability scores (Natale, 1975) to be predictive of degree of convergence, but studies on individual differences in convergence have not used retesting to establish individual consistency in convergence across tasks. Individual tendencies in convergence may exist but be outweighed by stronger differences such as the nature of each particular conversation. In that case, convergent tendencies of conversations may be more apparent than convergent tendencies of individuals. Such findings would support accounts of social factors mediating convergence (e.g. Giles et al., 1973; Babel, 2009; Pardo et al., 2012).

The methods used in Study 1 could not detect convergence patterns particular to each conversa-
tion, as each conversation provides only two data points per measure, one for each speaker. In order to evaluate both the individual properties of the conversation and the degree to which a particular conversation is convergent, we grouped together all the individual measures (which were already standardized). The six models in Study 1 each have a mixed effects model predicting the performance of the speakers based on their own baselines and the baselines of their interlocutors along a single measure; Study 2 tests convergence of each speaker across all measures. This provides up to 12 data points per conversation (6 measures for each speaker); not all measures were available for all conversations, given exclusion procedures based on sound quality and the usage of each characteristic.

The regression used each speaker’s performance in a conversation as the dependent variable (once for each measure). The speaker’s baseline in other conversations and the interlocutor’s baseline in other conversations were used as the two main fixed predictors. The interlocutor’s baseline in other conversation is the variable we identify with convergence: The more speakers’ productions are predicted by their interlocutor’s baseline, the more they are converging. These methods are the same used for the six individual measures in Study 1 except that each speaker in a conversation is repeated for every measure. The main difference is in the structure of the random effects. The characteristics of the model are outlined below.

Several of the possible random intercepts would be meaningless in the context of the current study. For instance, it is meaningful to expect per-interlocutor intercepts to differ within each measure, but measure-agnostic per-interlocutor intercepts would be meaningless, as that modeling choice would group together measures of very different nature. In contrast, measure-agnostic random slopes for convergence abstract over a putative single behavior, “participating in convergence,” and should therefore be included in the model. The modeling choices are listed below, and summarized in Table 6.

**Measure**  The model contains the per-measure random slopes, one for the speaker’s baseline and one for the interlocutor’s baseline. The two slopes are meant to account for the differences observed in speakers’ consistency and convergence in different measures, as observed in Study 1. Per-measure intercept was not included because in each of the individual models in Study 1, the
intercept was not significantly different than zero (the independent and dependent variables were all scaled).

**Conversation** We included a measure-agnostic random slope for convergence, and a per-measure random intercept. We did not include a measure-agnostic random intercept, as explained above. Within-measure intercepts are meant to explain the variance captured by per-conversation intercepts in Study 1. The random slope per conversation is one of four main variables of interest: If particular conversations elicit higher or lower degrees of convergence across multiple measures, this variable would significantly contribute to the model.

**Speaker** Despite its ultimate exclusion in Study 1 results, we included a within-measure random intercept per speaker. We also included a within-measure per speaker random slope for the interlocutors’ baseline to model within-measure individual differences in convergence, as well as a measure-agnostic random slope for the interlocutors’ baseline to model an overall tendency to converge. If some speakers are overall more or less convergent than others (“followers”), the slope would contribute to the model, making it better than a minimally different model that does not contain that slope. This slope is the second main variable of interest.

**Interlocutor** Interlocutor factors parallel the per-speaker random structure. We included a within-measure random intercept per-interlocutor, which mirrors the per-interlocutor random intercepts in Study 1 models. We also included both a within-measure and measure-agnostic random slopes for convergence. The former mirrors the per-interlocutor random slope for convergence in Study 1, and the latter models the possibility that some interlocutors elicit increased or decreased convergence across multiple measures (“leaders”). This is the third main variable of interest.

**Topic** For topic, we used a within-measure random intercept, which mirrors the per-topic intercept in the Study 1 models.
Table 6: The lme4 formula used to fit the model in Study 2

<table>
<thead>
<tr>
<th>lme4 syntax</th>
<th>explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>performance ~ 1 + speaker.baseline</td>
<td>Speaker’s performance in a conversation</td>
</tr>
<tr>
<td>+ interlocutor.baseline</td>
<td>Speaker’s baseline performance for the measure</td>
</tr>
<tr>
<td>+ (0 + speaker.baseline</td>
<td>study)</td>
</tr>
<tr>
<td>+ (0 + interlocutor.baseline</td>
<td>study:speaker)</td>
</tr>
<tr>
<td>+ (0 + interlocutor.baseline</td>
<td>study:interlocutor)</td>
</tr>
<tr>
<td>+ (1</td>
<td>study:topic)</td>
</tr>
<tr>
<td>+ (1</td>
<td>study:speaker)</td>
</tr>
<tr>
<td>+ (1</td>
<td>study:interlocutor)</td>
</tr>
<tr>
<td>+ (0 + interlocutor.baseline</td>
<td>speaker)</td>
</tr>
<tr>
<td>+ (0 + interlocutor.baseline</td>
<td>interlocutor)</td>
</tr>
<tr>
<td>+ (0 + interlocutor.baseline</td>
<td>conversation)</td>
</tr>
</tbody>
</table>

4.2 Results and discussion

As expected, both the speakers’ own baseline and their interlocutors’ baseline performance were significant predictors of the speakers’ performance ($\beta=0.711, SE=0.079, df=5, t=8.95, p=0.00029; \beta=0.0471, SE=0.011, df=5, t=4.45, p=0.00596$, respectively), though the former had a greater effect than the latter.

The results for the random effects are summarized in Table 7.

Table 7: Summary of Study 2 model random effects

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>Model comparison p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per-measure and conversation intercept</td>
<td>0.2473518</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Per-measure and interlocutor slope for convergence</td>
<td>0.0143233</td>
<td>0.86616</td>
</tr>
<tr>
<td>Per-measure and interlocutor intercept</td>
<td>0.0714960</td>
<td>0.00053</td>
</tr>
<tr>
<td>Per-measure and speaker slope for convergence</td>
<td>0.0459992</td>
<td>0.04416</td>
</tr>
<tr>
<td>Per-measure and speaker intercept</td>
<td>0.0000000</td>
<td>1.00000</td>
</tr>
<tr>
<td>Conversation slope for convergence</td>
<td>0.0000315</td>
<td>1.00000</td>
</tr>
<tr>
<td>Interlocutor slope for convergence</td>
<td>0.0453861</td>
<td>0.00160</td>
</tr>
<tr>
<td>Speaker slope for convergence</td>
<td>0.0000000</td>
<td>1.00000</td>
</tr>
<tr>
<td>Per-measure and topic intercept</td>
<td>0.2010929</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Per-measure slope for convergence</td>
<td>0.0226591</td>
<td>0.00123</td>
</tr>
<tr>
<td>Per-measure slope for consistency</td>
<td>0.1942646</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Among the variables of interest, only per-interlocutor slope for convergence significantly improved the model (SD=0.0454 \( p=0.0016 \)). We found no evidence for per-speaker slope for convergence (SD=0 \( p=1 \)), suggesting that the effect of “natural leaders” in convergence is easier to find than the effect of “natural followers.” Similarly, we found no evidence for per-conversation slope for convergence (SD=0.000315 \( p=1 \)).

As expected, there were significant differences in per-measure slope for consistency and per-measure slope for convergence (SD=0.194 \( p=<0.0001 \); SD=0.0227 \( p=0.0012 \), respectively). Similarly, as in Study 1, there were within-measure per-conversation differences and within-measure per-topic differences (SD=0.247 \( p=<0.0001 \); SD=0.201 \( p=<0.0001 \)). Unlike Study 1, the within-measure per-interlocutor differences did improve the model (SD=0.0715 \( p=0.00053 \)), as did the within-measure per-speaker slope for convergence (SD=0.046 \( p=0.044 \). Both differences could have arisen because the modeling of the fixed effects is not identical across measures. No other random effect had a significant effect on the model.

Some of the per-interlocutor consistency can be explained by the social status or likeability of the interlocutor; speakers converge more to more socially desirable interlocutors (Gregory and Webster, 1996; Hwang and Chun, 2018) and to interlocutors whom they like more (Yu et al., 2011; Sanker, 2015). However, our study was not specifically designed to test what characteristics of the interlocutor contributed to the differences in how much convergence they elicited, so that issue is left to future investigation. In contrast, per-speaker convergence slope did not improve the model. Given that the method was sensitive enough to find support for per-interlocutor tendencies in convergence, this null result is less likely to be due to insufficient power. Individual propensity to converge was not strong enough for our model to detect, providing little support for the view of convergence as an automatic tendency. The results provide striking support for the argument that variation in degree of convergence reflects properties of the interaction more than characteristics of the individual. Notably, per-interlocutor convergence is not a form of individual differences in the narrow sense. Interlocutors who elicit greater convergence have characteristics that make others converge to them, not internal tendencies for convergence. This result supports a socially-mediated explanation of convergence.
5 General Discussion

Our study addresses two main questions: is convergence a unified phenomenon, reflected similarly across measures, and are there intrinsic characteristics of speakers or interlocutors that create tendencies in convergence that differ across individuals. Both of these issues fit into a larger question of how convergence occurs and what factors mediate it.

Within our studies, convergence was evident to some degree in all measures, though with large differences in strength. Convergence is often explained as the result of changes in activation levels, both in lexical and syntactic priming and in exemplar accounts of phonetic convergence; our results are consistent with this type of model. Convergence in syntactic, lexical, and behavioral measures has been attributed to an intrinsic link between perception and production (e.g. Chartrand and Bargh 1999; Dijksterhuis and Bargh 2001), such that input automatically produces an activation in production. Repeated activation produces cumulative priming; people access lexical items and syntactic constructions more rapidly when they have recently been activated based on hearing them (Branigan et al., 2007), with stronger effects when the particular form or construction has been heard more times (Kaschak et al., 2011; Oben and Brône, 2016). There is similar evidence for a perception-production link in phonetics (cf. Motor Theory, Liberman and Mattingly 1985; Direct Realism, Fowler 1986), supported by neural evidence (e.g. Rizzolatti and Arbib 1998; Fadiga et al. 2002). Based on this link, convergence has been proposed as evidence for episodic memory (e.g. Goldinger 1998); acoustic details associated with phonological categories can have different weighting (e.g. Pierrehumbert 2002), based on which characteristics are salient given the language and the task (Johnson, 1997).

If convergence in different characteristics is driven by the same sort of perception-production link, then there should be a correlation in convergence in different measures, across conversations (Weise and Levitan, 2018). However, previous work has found such correlations to be weak (Pardo et al., 2012; Sanker, 2015; Weise and Levitan, 2018). Our study similarly finds little relationship between convergence in different measures across conversations or across speakers, though we do find per-interlocutor tendencies that are present across measures (see Section 5.2), which indicates that at least some of the variation in convergence aligns across measures.
Communication Accommodation Theory proposes that speakers use convergence and divergence to manipulate social distance from interlocutors in relation to membership in the same or different groups (Giles et al., 1973, 1991). However, socially mediated convergence does not necessarily need to be done consciously. Priming effects are actually stronger among individuals with declarative memory deficits than among others (Heyselaar et al., 2017) and imitation behaviors are also stronger in individuals with prefrontal lobe damage (Luria 1973:200-202). The presence of strong interlocutor influences within our data suggests that convergence is socially mediated, even if the mechanism driving convergence is not inherently social. Social factors in convergence may result from greater weight given to input from speakers who are viewed more positively. People tend to view their own attributes positively and thus also positively evaluate those attributes in others (Stalling, 1970; Byrne, 1971), which also provides a possible explanation for why speakers are more positively perceived when they converge (e.g. Dijksterhuis and Bargh 2001; Giles et al. 1973).

5.1 Weak Individual Tendencies in Convergence

Within our data, there was only weak consistency in convergent behavior from a given individual in each measure and no significant contribution of individual tendencies to predicting convergence across measures. On the other hand, the interlocutor was a significant predictor of degree of convergence in our data, so the lack of evidence for individual tendencies in convergence is unlikely to result from inadequate sample size, and seems to indicate that individuals do not have strong tendencies in convergence that are apparent across conversations or across measures.

Individual differences in perception have been demonstrated previously, and can have high test-retest reliability for the same task (e.g. Johnson et al. 1987) or tasks testing similar linguistic biases (e.g. Kong and Edwards 2016; Ishida et al. 2016). Individual differences in perception are often correlated between closely related tests, but correlations across less closely related tests tend to be weak and the differences do not correlate well with performance in most cognitive tests (e.g. Surprenant and Watson 2001; Kidd et al. 2007), though some individual differences in perception of linguistic stimuli correlate with cognitive measures such as Autism-Spectrum Quotient (Yu, 2013; Stewart and Ota,
There is more limited evidence for individual consistency in convergence. Some characteristics of the speaker are predictive of degree of convergence, for example, Autism Quotient (Yu et al., 2013), social desirability scores (Natale, 1975), social network size (Lev-Ari, 2018), and tendency to compromise (Weatherholtz et al., 2014). However, few convergence studies include re-testing to demonstrate that individuals are consistent in these differences, and the size of the effect is often small. There is some evidence for individual tendencies in convergence within a given measure between instances of same task or similar tasks, both in shadowing the same stimuli (Tamminga et al., 2018a) and in conversations with the same topic and different partners or different topics and the same partner (Sanker, 2015); evidence for tendencies across different tasks is much weaker (Pardo et al., 2018). Individual tendencies in convergence might be more apparent with more constrained conditions, because contextual factors like interlocutor and conversational topic are less likely to have an effect. The results of our study, in which each individual conversed with several different interlocutors on different topics, provide little evidence for individual tendencies in convergence; this result does not necessarily indicate that individuals lack such tendencies, but rather that such effects are small relative to other factors.

While some weak individual tendencies in convergence may exist, they are largely specific to the particular characteristic. Little previous work has tested correlations across measures; existing work that has looked for them has generally found no correlations (Bilous and Krauss, 1988; Pardo et al., 2012; Sanker, 2015; Weise and Levitan, 2018), with the occasional correlation that can probably be attributed to physical relationships between measures (Cohen Priva and Sanker, 2018) or repeated measures producing a false positive (Sanker, 2015; Rahimi et al., 2017). In studies with a relatively small number of participants, a null result for individual tendencies can be difficult to interpret; there has been no previous large-scale comparison across measures, controlling for effects of particular conversations. Our study, on a corpus with a large number of speakers, provides clearer evidence for a lack of individual convergent tendencies that hold across characteristics. This lack of relationship suggests that individual variability in convergence cannot reflect different processing styles at a broad
level (cf. Yu 2013; Yu et al. 2013). However, this does not mean that variation across participants is capturing nothing about individual tendencies. It is possible that there are individual differences that are just for processing very specific details (e.g. Hazan and Rosen 1991; Roy et al. 2017).

Convergence differs from measures of perception because it can also be influenced by individual differences in production and in the perception-production link, which may be why speakers have more limited individual tendencies in convergence than in perception. Some studies have found a correlation between patterns of perception and production (e.g. Zellou 2017; Perkell et al. 2004), though others have found a lack of significant correlation (e.g. Grosvald and Corina 2012; Schertz et al. 2015). That is, sometimes speakers who produce a pattern in their own speech are more likely to interpret acoustic stimuli as reflecting the same process, but results vary. Some consistency between perception and production patterns may be due to dialectal differences among the subjects (e.g. Harrington et al. 2008). The limitation in extending individual differences in perception to corresponding differences in production seriously limits the potential explanatory value of individual differences in propagating sound change.

5.2 **Stronger Interlocutor-specific Tendencies in Convergence**

The results of our study suggest that variation in convergence results more from aspects of the social contexts than internal tendencies of the speakers. This is consistent with previous work demonstrating that degree of convergence is predicted by actual or perceived characteristics of the interlocutor or model talker, such as dialect (e.g. Drager et al. 2010; Babel 2010; Weatherholtz et al. 2014) and status (e.g. Gregory and Webster 1996; Bane et al. 2010), and the speaker’s opinion of the interlocutor or model talker (e.g. Babel 2010; Yu et al. 2011). There is substantial variation in convergence based on the particular model talker, which can be consistent across different listeners (Pardo et al., 2017), though it has not been previously demonstrated that the per-interlocutor tendencies would be consistent across different conversations or different recordings of the interlocutor. While our studies do not focus on the particular aspects of the interaction that drive this variation, we crucially demonstrate that interlocutor-specific variation in convergence is consistent across conversations and across
measures. That is, the factors that strengthen or weaken convergence to a particular interlocutor in one linguistic characteristic similarly affect other characteristics.

While individual speakers do not have strong tendencies of being more or less convergent, individual interlocutors do have strong tendencies of eliciting more or less convergence. Similar variation based on the interlocutor or model talker has previously been demonstrated with smaller sets of individuals (e.g. Pardo et al. 2017; Hwang and Chun 2018). This result may suggest that previous work that found consistency in convergence in conversations containing the same individual (Sanker, 2015) is actually capturing consistency in convergence elicited by each person as an interlocutor, given that convergence was measured as change in distance within the conversation, which does not provide information about who is responsible for the change.

In contrast, the particular conversation was not a significant predictor of convergence. Some previous work has found that degree of convergence is predicted by aspects of the particular relationship between interlocutors (e.g. Pardo et al. 2012; Bane et al. 2010; Sanker 2015). However, given that such work has often looked at each participant in a single interaction, effects of the interlocutors and effects of the particular conversation could not be separated out, except indirectly based on finding differences associated with measurements of characteristics of the speakers or the conversation, or experimental conditions manipulating aspects of the conversation.

Consistency of degree of convergence across measures as predicted by the interlocutor limits what mechanism might be driving it, as some linguistic characteristics are differently influenced by cognitive demands or limitations. Thus, it is somewhat less likely that the same social characteristics of an interlocutor are producing variation by independently influencing different linguistic measures in the same way, and more likely that consistency of convergence across measures and conversations with the same interlocutor reflects interspeaker alignment (cf. Pickering and Garrod 2004), behaving similarly across measures, with different interlocutors eliciting different degrees of alignment. Different level of exposure to different characteristics of the interlocutors speech (Kaschak et al., 2011; Oben and Brône, 2016) could produce the absolute differences in convergence across measures; having the same contextual influences does not require that the absolute degrees of convergence would be the
same across measures.

The consistency of convergence for a particular interlocutor but not for a particular individual also limits some of the possible explanations about how variation arises. The variation produced by situational factors must be stronger than the individual tendencies, so any characteristic which is known to exhibit strong individual variation is unlikely to be driving differences in degree of convergence. For example, differences in attention are often proposed as a source of variation in degree of convergence, both due to characteristics of the individual (Yu et al., 2013; Namy et al., 2002) or aspects of the interaction, for example, listeners paying more attention to people they like (Dijksterhuis and Bargh, 2001) or paying more attention to speech addressed to them than to ambient speech (Brangan et al., 2007). However, individual differences in attention certainly exist (Fan et al., 2002; Tipper and Baylis, 1987), so the lack of evidence for individual tendencies in convergence suggests that these differences in attention are not a primary factor driving variation in convergence. Similar effects on convergence across measures also seem to oppose an attentional account, given that previous studies suggest different effects of attention in different measures. Heavy noun phrases exhibit stronger priming influences than lighter NPs (Francis et al., 2011), and syntactic priming is stronger among individuals with declarative memory deficits than among other speakers (Heyselaar et al., 2017). In contrast, distraction results in reduced phonetic convergence (Abel and Babel, 2017; Heath, 2017).

The variation in degree of convergence in with different interlocutors could result from differences in the weight given to input from the particular interlocutor; Chartrand and Bargh (1999) make a similar proposal, with more convergence resulting from greater perceptual engagement with the interlocutor. Normal conversational interaction may not produce enough variation in attention to result in variation in convergence, as speakers are reliably following what their interlocutors are saying. However, the impact of that input can depend on the social situation in which it is received. Greater convergence associated with more positive opinions of the interlocutor (Pardo et al., 2012; Sanker, 2015) or model talker (Babel, 2010; Yu et al., 2011) or higher status of the interlocutor (Gregory and Webster, 1996; Bane et al., 2010) is consistent with utterances from those speakers being given more weight based on their social importance, and also fits into explanations about constructing group iden-
tities (Giles et al., 1973, 1991). Stronger priming in goal oriented tasks than in conversation (Reitter et al., 2006) could also be consistent with greater local importance of exact phrasing in tasks requiring strict understanding. However, not all work has found this difference (cf. Pardo et al. 2018).

6 Conclusion

We present new data on convergence across measures and across speakers, using a large corpus of natural speech. Although there is variation across characteristics in convergence and in speakers’ consistency, different characteristics exhibited similar influence of particular conversations and particular interlocutors.

Variation across participants is not consistent across measures; to the extent that individual tendencies in convergence exist, they must be specific to particular characteristics. Within a characteristic, individual tendencies for convergence are weak relative to external factors such as the interlocutor. Thus, individual tendencies in convergence are unlikely to provide strong insights into language processing or the propagation of sound change.

On the other hand, variation across interlocutors was predictive across measures and across conversations; such differences have some potential to provide insights into the actuation of sound change. These results indicate the importance of situational factors in mediating convergence and moreover indicate that different linguistic characteristics are similarly influenced by situational factors, which provides a foundation for convergence studies that test conditioning factors using only a single measure of convergence.

References


