Causes and Consequences of Convergence

By

Jevon Scot Heath

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Keith Johnson, Chair
Professor Susanne Gahl
Professor Dacher Keltner

Spring 2017

Causes and Consequences of Convergence

Abstract

Causes and Consequences of Convergence

by

Jevon Scot Heath

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Keith Johnson, Chair

In speech convergence, people's speech becomes more like the speech they hear. Such convergence behavior has been observed along many domains of linguistic structure and in many different situational contexts. Convergence has been argued to be socially motivated (Communication Accommodation Theory — Giles et al. 1991), and also to be an unconscious, resource-free process (Interactive Alignment Theory — Pickering & Garrod 2004). This dissertation presents an alternative approach in which convergence is not a discrete process in itself; rather, convergence behavior is the consequence of episodic storage and recall, moderated by attention.

The first chapter of this dissertation consists of an elaboration of this approach, called the categorization schema account. In this approach, episodic storage is constrained by the categorization schemata that are currently active, and categories are only active when attention is paid to those categories' defining features. Convergence across disparate domains of linguistic structure is then an empirical pattern that falls out naturally from the assumption that multiple representations of the same input are stored separately and recalled independently. In consequence, speakers may converge to different domains of linguistic structure at different rates, depending on which domains have their attention.

The two subsequent chapters report the results of a pair of studies designed to examine predictions made by the categorization schema account. A Mechanical Turk experiment, discussed in Chapter 2, failed to find a significant difference between convergence to words and convergence to pseudowords. In a dyadic game task experiment comparing convergence rates across levels of linguistic structure, discussed in Chapter 3, participants exhibited different patterns of convergence to phonetic features on the one hand, and to lexical and syntactic features on the other hand. Additionally, participants who self-reported a greater degree of personal autonomy in this experiment exhibited less convergence behavior across domains.

Chapter 4 discusses the ramifications of these findings for theories of sound change, and reports the results of an experiment illustrating that accommodation can directly result in the appearance of new variants within an interaction, providing a possible pathway for the actuation of sound change.

to people

who do such amazing things

Table of Contents

List of Figures iv
List of Tables v

Chapter 1: Convergence and categories 1
1.0 Introduction 1
1.1 Convergence 2
1.1.1 Convergence is automatic 2
1.1.2 Convergence occurs at multiple levels of representation 4
1.1.3 Convergence is constrained by active categories 6
1.2 Categories 6
1.2.1 What creates categories? 6
1.2.2 How are episodes stored within categories? 8
1.2.3 What makes a category active? 8
1.2.4 What makes an episode an acceptable exemplar of a category? 9
1.3 The categorization schema account 10
1.4 This dissertation 12

Chapter 2: Episodic storage and recall 14
2.0 Introduction 14
2.1 Interacting levels of convergence 15
2.2 Episodic storage and social categories 16
2.3 Predictions 17
2.4 Experiment 17
2.4.1 Methodology 17
2.4.2 Results 19
2.4.3 Discussion 20
2.5 Conclusion 21

Chapter 3: Social effects on convergence 22
3.0 Introduction 22
3.1 Differential category activation 23
3.2 Predictions 26
3.3 Experiment 27
3.3.1 Methodology 27
3.3.2 Analysis 33
3.3.2.1 Analysis: Phonetic convergence 33
3.3.2.1.1 Convergence in stressed vowel duration 34
3.3.2.1.2 Convergence in vowel formants 37
3.3.2.2 Analysis: Lexical convergence 41
3.3.2.3 Analysis: Syntactic convergence 44
3.3.2.4 Analysis: Across levels 48
3.4 Discussion and conclusions 49

Chapter 4: Consequences of convergence **52**
4.0 Introduction 52
4.1 Non-faithful accommodation 52
4.2 The current study 54
4.3 Experiment 57
4.3.1 Methodology 57
4.3.2 Results across speakers 59
4.3.3 Results by speaker 60
4.3.4 Results by cue 63
4.4 Discussion 66

Chapter 5: Conclusion **70**
5.1 Convergence behavior is difficult—but possible—to isolate and measure 70
5.2 Convergence behavior is not specific to particular domains of linguistic structure 71
5.3 Convergence behavior is predicted by personal autonomy 71
5.4 Convergence behavior is a potential actuator of sound change 72

References **73**
Appendices **79**

List of Tables

Table 1.1: Meanings of *automatic* 4
Table 1.2: Candidate categories for the utterance "*Okay*." 5
Table 2.1: Vowel Euclidean distance from model talker in shadowing task 20
Table 2.2: Hypotheses under consideration for frequency-convergence interaction 21
Table 3.1: Differences between (A) *I would like a cookie* and (B) *Can I have a cookie?* 25
Table 3.2: Game experiment objects 29
Table 3.3: Percentage name agreement (from Snodgrass & Vanderwart 1980) 29
Table 3.4: Starting distribution of objects 30
Table 3.5: Requirements by round/player, Iteration 1 31
Table 3.6: Requirements by round/player, Iteration 2 32
Table 3.7: Categorical convergence paradigm 33
Table 3.8: Distribution of first/last thirty vowels 34
Table 3.9: Difference in vowel duration predicted by time point and vowel 36
Table 3.10: Difference-in-distance of F1/F2 in vowels 38
Table 3.11: Vowel F1/F2 distance model (no power predictors) 39
Table 3.12: Vowel F1/F2 distance predicted by community activism and autonomy 40
Table 3.13: Lexical convergence predicted by frequency of forms used 42
Table 3.14: Lexical convergence predicted by frequency and community activism/autonomy 43
Table 3.15: Syntactic convergence predicted by type of utterance used 46
Table 3.16: Convergence to utterance predicted by type of utterance and community activism/autonomy 47
Table 3.17: Correlation table of participant convergence rates across levels 47
Table 3.18: Convergence as a function of level of linguistic structure 49
Table 3.19: Hypotheses under consideration for game experiment 50
Table 4.1: Possible speaker adjustments for coincident cues across conditions (model with longer VOT and closure than speaker) 56
Table 4.2: Global changes across conditions 59
Table 4.3: Individual changes across conditions (vowel cues) 60
Table 4.4: Individual changes across conditions (word-medial stops) 61
Table 4.5: Random slopes by speaker for VOT across conditions 63
Table 4.6: Correlations of characteristics and changes across conditions 67

List of Figures

Figure 1.1: Categorizing an instance of the utterance "*Okay*." 10
Figure 3.1: Pictures of a bee/wasp and pan/saucer (from Snodgrass & Vanderwart 1980) 28
Figure 3.2: The game experiment in progress 30
Figure 3.3: Absolute difference in vowel duration by subject 35
Figure 3.4: Community activism & autonomy and lexical convergence 44
Figure 3.5: Community activism & autonomy and convergence to utterance type 48
Figure 4.1: Antagonistic accommodation leading to innovated variation 53
Figure 4.2: Example of stop duration measurement for the word ATTESTED 59
Figure 4.3: Global changes across conditions 60
Figure 4.4: Antagonistic accommodation (S25) and hyperconvergence (S35) in closure duration 62
Figure 4.5: Vowel F0 by speaker 64
Figure 4.6: Vowel duration by speaker 65
Figure 4.7: Stop closure duration by speaker 66

Acknowledgments

My hope is that everyone whom I should mention here already knows how important they have been to me, to my work, and to my happiness. I further hope that this isn't the first time I'm telling you how you've affected my life. If I haven't told you before, I truly apologize, and I hope I am doing so now.

Even before I knew linguistics was a field I knew there was something there that I loved. It took me a while to understand what it was that interested me, and I had a lot of help from a lot of people in getting to that point. Gerald Heath, Lucy Harville, Sarah Heath, Lupe Santana, Susan Kircos, Aron Bothman, Dug, Ben Harville, Andy Warner, Sam Liebhaber, Isabel Downs, Katelyn Gamson, Samara Weiss, Patricia Ruth, Jelena Teague, Ognjen Smiljanić, Michael Furlow, Kavin Scalf, Carol Genetti, Marianne Mithun, Matt Tentler, Matt Tucker, and Bob Wagoner: I am here because you showed me where I was going.

Literally everybody in the UC Berkeley Linguistics department should be named here, so I will call out those whose influence on me has been acute. Foremost, my cohort — Nicholas Baier, Erin Donnelly, Matthew Faytak, Joseph Giroux, Matthew Goss, John Merrill, Kelsey Neely, and Melanie Redeye — helped keep everything in my head that should be kept there. Thank you, Larry Hyman, Andrew Garrett, Line Mikkelsen, Sharon Inkelas, Peter Jenks, Keith Johnson, Susanne Gahl, and Susan Lin. Thank you, Justin Spence, Will Chang, Roslyn Burns, and Gregory Finley. Thank you, Kelsey Neely, Peter Jenks, El-Haji Malick Loum, and Vivian Wauters.

In coming to the conclusions I am presenting here I was guided and enlightened by discussions with many people, including Sharon Inkelas, Alan Yu, Jeff Mielke, Morgan Sonderegger, Molly Babel, James Kirby, Kuniko Nielsen, Susan Lin, Kevin McGowan, and Auburn Barron-Lutzross. I was ably assisted in collecting and analyzing my data by Christine Jiang, Julia Morse, Stella Gerdemann, Rebecca Hong, Julay Brooks, and Justin Knight. I had a wonderful and enthusiastic committee in Keith Johnson, Susanne Gahl, and Dacher Keltner. I also need to contribute mine to the manifold thanks due Ronald Sprouse, Paula Floro, and Belén Flores.

Nico gets his own paragraph, for some reason.

During the preparation and writing of this document I found comfort, advice, and sanity in my interactions with wonderful friends. I so greatly appreciate my conversations and diversions with Marika Kuzma, Alex Dougal, Randy O'Connor, Vivian Wauters, Marc Juberg, Emily Cibelli, Hannah Sande, Andrew Cheng, Kenny Baclawski, Nik Rolle, Marcus Ewert, Elise Stickles, Florian Lionnet, Matt Goss, Keith Weissglass, Andy Warner, Kathy Guis, and Zachary Seldon. Thanks to Emily Remirez for putting my enamel in danger. I especially want to acknowledge Hannah Sande, Gregory Finley, and Stephanie Farmer, for our weekly support group meetings. Even though it's been a while.

And… Thank you.

And finally… Thank *you*.

Chapter 1: Convergence and categories

1.0 Introduction

People change the way they speak after hearing speech, largely independent of the content of the speech they hear. Studies have repeatedly and robustly demonstrated instances of this phenomenon, at many different levels of linguistic structure as well as in different features within the same level. Different disciplines have given an assortment of names and accounts for manifestations of this phenomenon at these different levels, including *entrainment*[1], *syntactic persistence*[2], *accent convergence*[3], *phonetic convergence*[4], *imitation*[5], *synchrony*[6], *co-ordination*[7], *adaptation*[8], *accommodation*[9], *alignment*[10], and *phonetic imitation*[11]. These demonstrations all share a tendency toward *convergence*: people's speech tends to be more like the speech they hear. However, it has largely been the case that each level of linguistic structure gets its own explanation for occurrences of this phenomenon within it. The aim of this dissertation is to move toward an account of convergence in which these various approaches can be reconciled.

Conceptually speaking, there are two main ways to converge, corresponding to Marr's (1982) distinction between the computational and representational levels of analysis. These are convergence toward *what is being done*, and toward *the way something is being done*. *What is being done* consists of the goal-defined words or actions that an individual utters or performs; for example, nodding one's head versus shaking one's head, or saying *Yes* versus saying *No*. *The way something is being done* consists of the specific details of these larger-scale words and actions; for example, raising one's hand by first rotating the shoulder and then flexing the elbow, versus flexing the elbow first; or saying the word *Yes* with a slowly rising intonation and a lengthened vowel, versus with an abrupt falling intonation and a short vowel. The difference between, e.g., nodding one's head and saying *Yes* may fall under either paradigm, depending on whether it is the goal or the way of achieving that goal that is more important for the current context.

This distinction between *what is being done* and *the way something is being done* mirrors an important distinction in research on memory and language production, between procedural and episodic memory. Procedural memory includes unconscious memories of procedures; episodic memory includes memories of specific events. In speaking, syntactic processing has been argued to draw on procedural memory (e.g. Ferreira et al. 2008), whereas phonetic detail within lexical retrieval draws on episodic memory (e.g. Goldinger 1998).

---

1. Levitan & Hirschberg 2011.
2. Bock 1986.
3. Bourhis & Giles 1977.
4. Pardo 2006.
5. Goldinger 1998.
6. Webb 1970.
7. Branigan et al. 2000.
8. Gregory & Hoyt 1982.
9. Giles et al. 1973.
10. Pickering and Garrod 2004.
11. Babel 2012.

The central idea I am pursuing is that convergence toward *what is being done* is a consequence of procedural memory, which activates procedures that a person has formed through their experience with language. Active procedures are informed by active categories, which are held in working memory. Convergence toward *the way something is being done* is a consequence of episodic memory, which activates the available categories that a person has determined, consciously or unconsciously, apply to the current context. Multiple representations are stored separately and can be recalled independently. This effective difference in the use of episodic versus procedural memory is in line with previous approaches such as Ullman's Declarative/Procedural model (2001, 2004), in which the mental lexicon depends on declarative memory (which includes episodic memory) while the mental grammar relies on procedural memory.

The approach presented here assumes that convergence at various levels of linguistic representation occurs in an equivalent way, subject to the formal constraints of the levels. As such, I will have to reconcile accounts of convergence that have focused on particular realizations of this general phenomenon. The current account stands in opposition to Communication Accommodation Theory (Giles et al. 1991), which claims that convergence is an intentional process. While convergence of vowel formants is phonetically and socially selective, it is automatic within those restrictions (Babel 2009). Conversely, Pickering and Garrod (2004, 2013) describe an Interactive Alignment Theory of dialogue, in which alignment, defined as "when interlocutors share the same representation at some level" (2004:172), is a mostly unconscious, resource-free process that cannot be switched off: "The activation of a representation in one interlocutor leads to the activation of the matching representation in the other interlocutor directly" (2004:9). In Pickering and Garrod's model, alignment results from "implicit common ground", which is built up between interlocutors over the course of a dialogue. Individuals in conversation align their situation models in such a manner that their production and comprehension systems interact at all levels of linguistic processing. This alignment is driven by priming, in that every utterance automatically primes all associated representations for both speakers. The account I am proposing agrees with this account in that speakers do not have control over the activation of speech procedures. However, speakers are able to manipulate which categories they have active and how they inform those procedures, based on selective attention. Pickering and Garrod's "common ground" is not then a mechanism in itself; it is instead a consequence of social effort that interlocutors put into their interaction. Additionally, the account I am proposing does not require that interlocutors share representations in order to evince convergence, although it does allow for this possibility.

1.1 Convergence

1.1.1 Convergence is automatic

There is a general consensus that convergence is default behavior (Bourhis & Giles 1977, Tannen 1987, Delvaux & Soquet 2007, Babel 2009, 2012, *inter alia*). Some direct evidence for the lack of speaker control over convergence comes from Lewandowski (2012), who found that native English speakers could not avoid converging toward their native German-speaking interlocutors even when explicitly instructed not to alter their pronunciation to accommodate their interlocutors' non-native accents. However, when speakers *try* to imitate, they do worse at it. Pardo

et al. (2010) looked at the interaction of conversational role and explicit instruction in a map task, where one participant gave another directions on a map. They found increased convergence when the information giver was instructed to explicitly imitate their interlocutor; however, when the receiver was instructed to imitate, participants ended up diverging. Episodic models of speech production (Pierrehumbert 2001) provide a good account of this lack of control: as instances of a particular category are perceived, they become part of the definition of that category; and subsequent productions of instances of that category are influenced by the category's new definition.

Word frequency effects also support an episodic theory of convergence. Several studies of convergence have found that word frequency inversely correlates with the degree of convergence evinced by speakers (Goldinger 1998; Goldinger & Azuma 2004; Nielsen 2011, 2014). In an episodic theory of lexical representation (Goldinger 1998), this correlation is explained by the activation of word-level categories. Any given token of a low-frequency word constitutes a greater proportion of a speaker's experience with that word, such that a new token of a low-frequency word will adjust the speaker's expectations about that word's distribution to a greater extent. As such, the activation of word-level categories will lead to this kind of word frequency effect. The same line of reasoning applies to findings of Goldinger (1998) and Pardo (2006). In the former case, Goldinger found greater convergence in a shadowing task with more repetitions of a word. In the latter, Pardo found greater convergence between interlocutors at later time points within a conversational setting. As the episodes within the experimental setting constitute a larger proportion of the active word-level categories, the speaker's production target will be more heavily weighted toward those episodes.

However, syntactic priming is likely not provoked by episodic memory. Syntactic priming is resistant to forgetting, whereas declarative memory (which includes episodic memory) shows decay over time and with intervening material (Bock et al. 2007, cited in Ferreira et al. 2008). As Bock & Griffin (2000) showed, sentences that are explicitly remembered are no more likely to cause syntactic persistence, and vice versa. Ferreira et al. (2008) found that patients with anterograde amnesia exhibited syntactic persistence effects at an equivalent rate to matched control subjects, despite being worse at recognizing that they had heard the very sentences that induced priming. They argued that syntactic priming is driven by procedural, rather than episodic, memory.

We can characterize syntactic priming as convergence toward *what is being done*, if we treat a syntactic structure as a tool for situating concepts in relation to one another. Other non-linguistic behavioral imitation patterns have demonstrated convergence to *what is being done* in this way, independently of *the way something is being done*. For instance, Bekkering et al. (2003) looked imitation of hand gestures in preschoolers in reaching for various objects. They found that the children faithfully imitated the goal of the gestures (*e.g.,* touching a particular dot on the table) over 90% of the time, but faithfully imitated which hand was being used to make the gesture only 76% of the time when doing so meant crossing their body.

So if convergence is default behavior, is it necessarily automatic? Delvaux & Soquet (2007) describe phonetic imitation as an "automatic" process: "[...] evidence for a general tendency for spontaneous and automatic imitation of the way of speaking of the ambient language, regardless of the complex history and nature of the social relationships that may exist between 2 interacting speakers" (2007:148). But what is meant by "automatic"? Schneider & Shiffrin (1977) define a process as automatic within a system when it occurs without the subject's control, without

requiring the subject's attention, and without stressing the system's capacity. Babel (2009) draws a distinction between two meanings of the word: "[...] phonetic imitation is not automatic in terms of occurring all the time, but indeed automatic in terms of happening subconsciously. That is, the social factors that mediate the imitation process are not explicit social choices, but implicit socio-cognitive biases" (2009:2). I think that there are many meanings of this word that may be intended by different writers at different times:

Table 1.1: Meanings of *automatic*

- *systematic* — the process always takes place, whether it has an effect or not
- *unconscious* — the process takes place without conscious direction
- *inevitable* — the process always ends up with the same result, regardless of input
- *reflexive* — the process always takes place given the appropriate trigger conditions
- *inviolable* — once begun, the process always takes place from start to finish
- *consequent* — the process is not autonomous, rather it is a consequence of other factors

Schneider & Shiffrin's (1977) definition of automatic processes then includes *unconscious and reflexive* components. Pickering and Garrod's (2004) Interactive Alignment Theory discusses dialogue as automatic, meaning that it is *reflexive* and *consequent*. Given the abundant evidence that it is affected by social factors (Bourhis & Giles 1977; Abrego-Collier et al. 2011; Babel 2010, 2012), imitation is clearly not *inevitable*. Indeed, the influence of social factors is part of Delvaux & Soquet's claim: "[U]nless hindered by higher-order sociopsychological factors (e.g. deliberate will to dissociate from a particular social group, or to distance oneself from a specific individual), speakers automatically tend to adjust their phonetic realisations to ambient speech" (2007:146). This quotation also indicates that Delvaux & Soquet, at least, do not use "automatic" to mean *inviolable*, as they explicitly allow for intervention from outside processes. However, any of the other definitions are still in play.

For my own discussion of automaticity in convergence, I am using "automatic" to mean *reflexive*. I do not intend to imply that convergence always takes place, nor that it is a discrete process, nor that it always results in the same outcome. While the categorization account put forward here does treat convergence as *consequent* of exemplar storage and retrieval, I do not use the term "automatic" to refer to this fact. Likewise, I do not intend for the unconscious nature of convergence to be at issue here.[12]

1.1.2 Convergence occurs at multiple levels of representation

As Pierrehumbert (2001:4) says: "It is important to note that the same remembered tokens may be simultaneously subject to more than one categorization scheme, under such a model. For example, a recollection of the phrase *Supper's ready!* could be labelled as "Mom" and "female speech", in addition to exemplifying the words and phonemes in the phrase." Table 1.2 lists some possible convergence schemata for the utterance "*Okay.*"

12. Bourhis & Giles's (1977) irate Welsh student notwithstanding.

It seems sensible to take as given that utterances are perceived and processed at multiple levels at once. There is considerable evidence that speakers converge along multiple levels as well; although most convergence studies have looked at convergence in one level of linguistic structure at a time, there have been many independent efforts to test convergence at many levels. Some of these studies are summarized in Chapter 2.

Table 1.2: Candidate categories for the utterance "*Okay.*"

the word *okay*
the utterance *Okay.*
the phoneme /oʊ/ | /k/ | /eɪ/
diphthongs
aspirated voiceless stops
the syllable /ʔoʊ/ | /keɪ/
stressed syllables
unstressed syllables
speech in English
speech in California English
speech in Chicano English
utterances with a slightly rising intonation
speech by a friend
female speech
Maria's speech
Maria's speech when she's angry at me
speech in living rooms
speech in my living room
speech directed at me
the conversation I had with Maria on June 6, 2006 in the early afternoon
acceptance of a proposition
angry speech
passive-aggressive speech
speech ending a conversation

One study examining convergence at multiple levels at the same time is a shadowing study by Nye and Fowler (2003), in which participants shadowed strings of phonemes that approximated English to varying degrees. The phoneme strings were produced by a trained phonetician, but still contained coarticulatory cues. The authors found greater convergence for more arbitrary strings of phonemes than for phonotactically probable sequences; that is, they found differing convergence at phonetic and phonological levels of representation depending on which was more informative for the situation.

1.1.3 Convergence is constrained by active categories

However, convergence does not always happen. Some studies have indicated that convergence is moderated by speaker attention. Goldinger (2013) found that imitation increases when the named objects have competitors with similar names or appearances, indicating that speakers converge more when they have to pay more attention in order to discern specific differences. Babel et al. (2014) conducted a shadowing task in which participants shadowed words produced by eight model talkers in eight successive blocks. The model talkers consisted of the voices of talkers of each sex from a collection of 60 voices that were rated as most and least attractive and most and least typical in a prior experiment. Babel et al. found that speakers of both sexes converged more toward the attractive and atypical voices of both sexes than toward the unattractive and typical voices of both sexes, and imputed this finding to speakers' social preferences. However, a study by Sui & Liu (2009) found that attractive faces were more effective distractors in a spatial cuing task than unattractive faces, suggesting that facial beauty competes for at least one kind of attention. As such, Babel et al.'s findings may also be consistent with an interpretation of attention, rather than preference, moderating convergence.

The possible influence of attention on convergence points to the intervention of working memory. Working memory capacity reflects the ability to keep a representation active, even or especially despite distractions (Engle et al. 1999; Engle 2002). All categories are not active at all times; the number of categories that are simultaneously active is restricted by working memory constraints.

Accommodation studies have shown convergence in nearly every feature that has been investigated. Under the current account, the intercession of attention on convergence may explain why this is the case. In designing an experiment intended to target particular features, experimenters end up highlighting those features such that the best strategy for processing the input within the context of that experiment is to draw on categories that differentiate between those features. Concurrently, features that are not being targeted are not systematically controlled, such that statistical analyses of those uncontrolled features are much less likely to be carried out in a manner that researchers are confident in. This tendency also unfortunately makes it difficult to assess whether my claim about differential convergence patterns involving episodic vs. procedural memory is consistent with the literature, as most studies have heretofore focused on one particular type of categorization at a time.

1.2 Categories

1.2.1 What creates categories?

One clear prediction of episodic models of convergence (Pierrehumbert 2001, Goldinger 1998, etc.) is that convergence behavior should be malleable, and should therefore change with age and experience. I expect different individuals to develop categories at different stages, in different orders, as their personal experience dictates. Categories of proper behavior will tend to be defined and informed by observed behavior. Differing hierarchies and relationships between categories may drive differing behavior in the present.

When a new situation is encountered, competent speakers initiate a new situational context-level category in anticipation of any difficulties that may be specific to the situation. The new category is formed from a template consisting of a previously existing context-level category that is anticipated to be similar in some way. This means that new contexts are approached with untested expectations already in place.

A consequence of templates for new contexts being built off of previous experience is that speakers with more linguistic experience in general will evince different behavior in convergence than less experienced speakers. In this vein, several studies have looked at age as a predictor of convergence (Kent 1979; Kuhl & Meltzoff 1996; Nielsen 2014). Their findings are coherent with an interpretation of accuracy in convergence increasing with experience across the lifespan. Kent (1979) looked at fidelity of imitation of 6-year-old children and adults to fifteen synthesized vowels, measuring the first three formants. Five of the synthesized vowels approximated English vowels; these were imitated by both groups, although the children showed greater variability. While both groups were less accurate in imitating the ten non-English vowels, adults were significantly more accurate than children.

In comparing the vowel productions of 12-, 16-, and 20-week-old infants imitating recorded adult productions of /aiu/, Kuhl & Meltzoff (1996) found that older infants had more discrete vowel categories in F1/F2 Euclidean space, as well as in Compact-Diffuse/Grave-Acute space.

Nielsen (2014) looked at VOT convergence in preschoolers, third graders, and (college-age) adults. Subjects were exposed to /p/ and tested on new and old /p/ as well as /k/, as in Nielsen (2011). Unlike the Kent (1979) and Kuhl & Meltzoff (1996) studies, Nielsen's study did not involve explicit instructions to imitate. Despite this, all three participant groups extended VOT in imitation. However, both groups of children extended VOT by twice as much as adults did.

Another partially-overlapping set of studies indicate that familiarity with a particular language variety predicts convergence toward speech of that variety. Studies by Chistovich et al. (1966) and Kent (1979) show that speakers converge more toward features that they already evince in their own speech. Chistovich et al. (1966) found that the native Russian speaker in their study converged more to synthesized vowels that were more like typical productions of Russian vowels; Kent (1979) found an analogous result in English. Mitterer & Ernestus (2008) found that Dutch speakers in general did not change whether they produced alveolar or uvular trills in accommodation to a model. (Although both types of trills are permissible in Dutch, most speakers only produce one or the other.) Kim, Horton & Bradlow (2011) found that same-dialect dyads of English or Korean native speakers exhibited greater convergence.

Taken together, the results of these studies support an episodic approach to speech processing, with more experienced speakers possessing a greater array of better-defined categories. These findings also fit into research in other areas of social development. For example, Garrod and Clark (1993) found that younger children were not able to avoid alignment in coordinate description schemes during a joint map task. This suggests that examining the accommodation processes of children at various stages of social development will potentially enable us to chart the order of acquisition of these categories.

Another possible explanation, however, is that it is not social mediation that children acquire, rather a greater range of experience more generally. As they grow older, children acquire more episodes for words, phonemes, contexts, and so on. In so doing, the effects of any new tokens will be less impactful.

1.2.2 How are episodes stored within categories?

Not every aspect of an episode is stored; only those aspects that pertain to the membership of activated categories are stored. This is a partial answer to the head-filling-up problem (Johnson 1997): if every detail of every exemplar were remembered, the memory demands would be impossibly oppressive. Some evidence for this approach lies in findings that people evince greater convergence when less processing can take place, either because of the novelty of received stimuli (Babel et al. 2014) or because of the short timeframe being investigated (Goldinger 1998). If some sort of filtering takes place before perceived stimuli are stored, people will converge more to stimuli that haven't been processed yet, as that filtering will not yet have taken place.

Novelty makes episode storage more likely. A shadowing study by Nye and Fowler (2003) indicates that different levels of accommodation can occur even within a shadowing task. In this study, participants shadowed strings of phonemes that approximated English to varying degrees. The phoneme strings were produced by a trained phonetician, but still contained coarticulatory cues. The authors found greater convergence for more arbitrary strings of phonemes than for phonotactically probable sequences. If we assume that more attention in general is paid to episodes that are unfamiliar, we predict this finding.

Once an episode has been stored in the appropriate categories, many details about it will be discarded as extraneous or redundant information. This is especially the case when those details are predictable, i.e. when the episode is closer to the prototype for the activated category. This is not to say that only one category is active at any given time; multiple levels of categories may be active to varying degrees, depending on the situation at hand. Which details are to be retained are specific to each category; different details about the same episode may be retained in different active categories. The number of active categories at any given time may be limited by working memory considerations.

This discarding of irrelevant detail to active categories is in accordance with findings that people evince greater convergence when closely shadowing segments (Chistovich et al. 1966) or words (Goldinger 1998) than when repeating them after a delay. Chistovich et al. played synthesized front vowel-like stimuli and had a listener shadow, mimic, and transcribe the stimuli in separate conditions. They found that delayed mimicking resulted in more categorical production, with F2 and F3 values of the produced vowels forming a clear delineation between two of the points along the continuum. In shadowing, the subject did show evidence of categorical production, but to a less pronounced degree. Similarly, Goldinger (1998) found that speakers were perceived as converging to the model talker to a greater degree when closely shadowing recorded words than when repeating them after a delay of a few seconds. Assuming that all retained detail informs subsequent productions, these findings support a theory in which some details are discarded, such that repetitions produced after information is discarded are not informed by the discarded information.

1.2.3 What makes a category active?

Different criteria for categorization are appropriate at different times. If you're trying to find a friend in a crowded bar by listening for their voice, you do not need to attune to the words they are saying; paying attention to intonation patterns, fundamental frequency, and other

individual-specific cues will be more useful. On the other hand, if you're trying to find a group of friends who are expecting you, it is a better strategy to listen for your name than for any individual voice.

The word *coat* in the sentence *Where's my coat?* is relevant to many different levels of category, among which are: word class (*Instance of a noun*); word (*Instance of the noun COAT*); and phoneme (*Instance of /k/*). Each of these categories is in opposition to or competition with other categories at the same level of linguistic structure (e.g., *Instance of a verb; Instance of the noun CAMERA; Instance of /g/).* I will refer to these separate levels of category as *categorization schemata*. A given categorization schema encompasses all of the competing categories at that level of structure. By activating an appropriate categorization schema, individuals are able to focus on the pertinent types of cues for a given situation.

While a tremendous number of categories are appropriate and useful in different circumstances, not all categories can be active at any given time. Because only a subset of categories are active at once, received speech that falls outside of the active categories is harder to process until the appropriate categories have been activated. I am making the claim that the process of determining which categories are appropriate for the given situation is an unconscious process, but one that is informed by contextual, experiential, and social factors. Once the appropriate categories are determined, the activation of those categorization schemata is reflexive. The selection of appropriate categories is a skill that speakers learn over their lifetime.

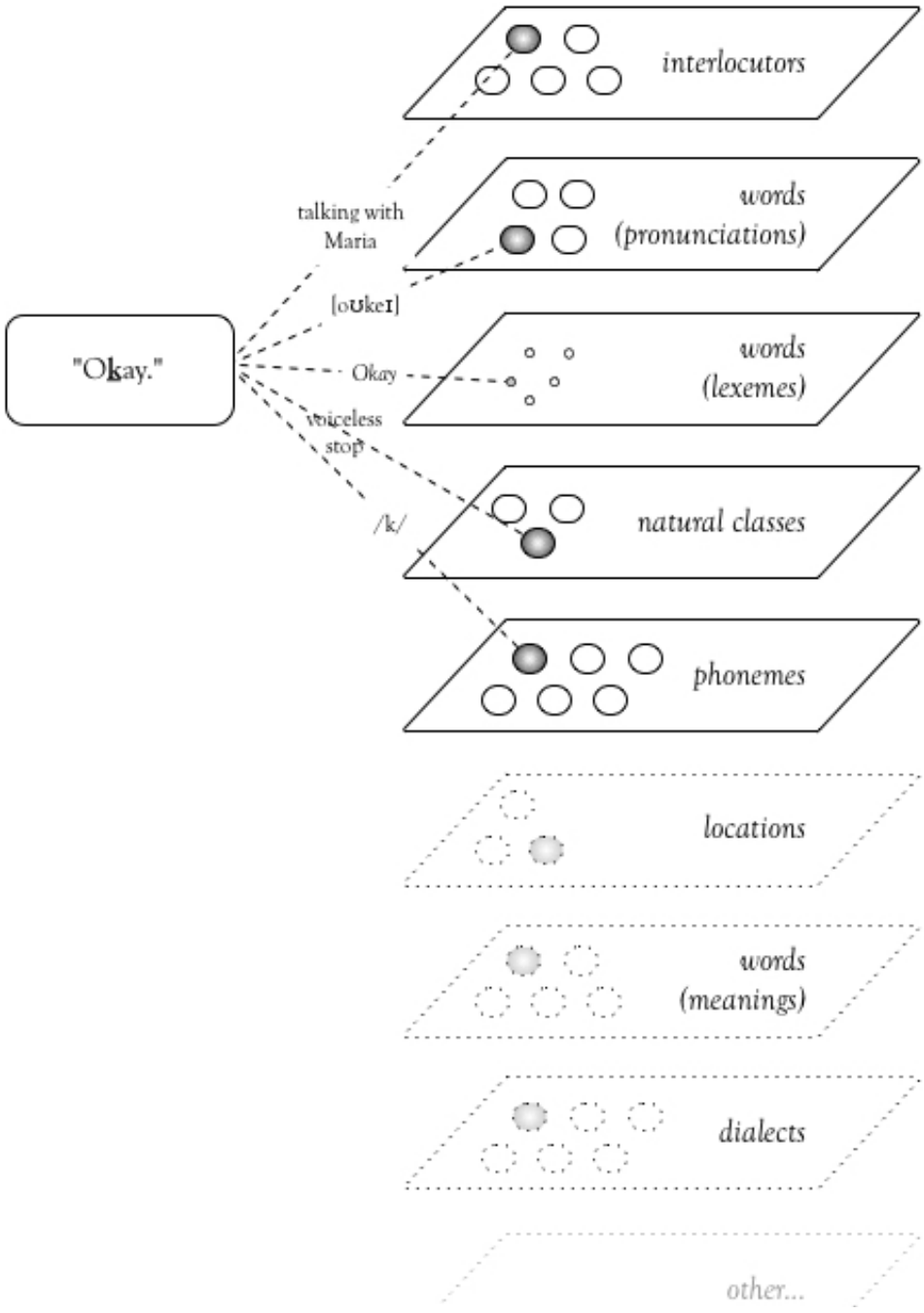
1.2.4 What makes an episode an acceptable exemplar of a category?

Social factors affect perception. This has been shown in studies of speaker normalization (Strand and Johnson 1996; Strand 1999), perception of diphthong quality (Niedzielski 1999), and dialect familiarity (Sumner and Samuel 2009), among other areas. Dijksterhuis and Bargh (2001) propose the concept of social perception: people perceive not only social actors' direct behavior, but also their own inferences about traits that actors have as well as stereotypes related to actors' group membership.

An episode is an acceptable member of an activated category if it is perceived to match the definition of that category. However, as episodes are not perceived veridically, it is possible that a given episode will not become an exemplar of an active category that it "should", due to e.g., it not being noticed that the episode in fact has a crucial defining characteristic of the category. Under the current proposal, this locus of potential mismatching is the entry point through which social factors may affect convergence: any inhibition of convergence is due to inhibition of matching perceived episodes to active categories.

Practice with a particular pattern of variation will improve accurate perception of that pattern. Sumner and Samuel (2009) found that experience with a dialect strongly affects a listener's ability to recognize words spoken in that dialect, and that region-general variants generally abet successful recognition. The authors conducted a series of experiments including participants unfamiliar with "r-less" varieties of American English, participants who actively spoke such a variety, and participants who were extensively familiar with r-less English but who generally spoke an "r-ful" variety. They found that hearing an r-less pronunciation primed listeners who were familiar with r-less English, but hearing an r-ful pronunciation primed all listeners.

Figure 1.1: Categorizing an instance of the utterance "*Okay.*"



1.3 The categorization schema account

Under the current account, people have many different ways of categorizing language, and only some of them are active at any given time. Any incoming speech is only processed by the currently active categorization schemata. A new episode updates the active categories that it fits, but does not update inactive categories that it fits. Furthermore, some categories do not store variation, as there is nothing to update. For example, the lexeme *okay* is composed of a strict string

of phonemes with a particular syntactic paradigm, and these relationships will not change. Figure 1.1 illustrates this account for the utterance *"Okay."*

As an example of the categorization schema account, I now lay out the way in which this theory addresses the results of Nielsen's (2011) phonetic imitation study. In this study, participants were exposed to model speech with extended voice onset time (VOT) exclusively in words beginning with /p/, and then recorded saying words beginning with both /p/ and /k/. Participants extended their VOT in words beginning with both phonemes, although new words with initial /p/ evinced a higher degree of VOT adjustment than words with initial /k/. Words with a lower lexical frequency showed a higher degree of VOT adjustment than more frequent words. In another condition, participants were exposed to model speech with *shortened* VOT; they did not evince convergence in this condition.

Interpreted through the proposed theory, this study involves the interrelation of five different categorization schemata: two different word-level categorization schemata, in which episodes of each word are categorized separately as instances of that word being represented (*what is being done*) and as instances of that word being pronounced (*the way something is being done*); a single phoneme-level categorization schema, in which episodes of /p/ are stored separately as instances of /p/; a phonological natural class-level categorization schema, in which episodes of voiceless stops are stored separately as instances of voiceless stops[13]; and a situational context-level categorization schema, in which episodes in this experiment are stored separately as instances of speech in this experiment. Every time a participant hears a word during the study, a memory of that episode is stored in the appropriate category in each of the active schemata. Procedural memory is only invoked in this study in recognizing and categorizing the words that are being said.

Suppose that the first word heard is "parrot" with an extended VOT for the initial stop. The stimulus as a whole is stored as an episode of experimental speech. Additionally, the stimulus is stored as an episode of the word *parrot*, and the mean VOT of episodes of the word *parrot* is increased slightly. The stimulus is also recognized as an instance of the word *parrot*, and so the lexical representation of the word is activated. The initial phoneme of the stimulus is stored as an episode of the phoneme /p/, and the mean VOT of episodes of the phoneme /p/ is increased slightly. Finally, the initial stop of the stimulus is stored as an episode of the class of voiceless stops, and the mean VOT of episodes of voiceless stops is increased slightly, although by less than the mean VOT of the phoneme /p/.

Suppose then that the second word heard is "pasture" with an extended VOT. The stimulus is stored in the same way as with "parrot", except that the word-level categories are different. The experimental category now contains episodes from different word-level categories that are unusual in the same way, namely that their initial stops have longer than average VOT. At this point, the speaker's word-level representation schema will likely deactivate, as the meanings, distributional information, etc. that are associated with representations at this level are not being used within the experimental context.

When the time comes to produce words in the post-task phase of the experiment, the participant will draw on all appropriate categorization schemata to produce the appropriate word, to the extent dictated by the situational context as it is understood. If the word to be produced is "parrot", the participant will draw on episodes from the "experimental setting" category of the

13. This categorization schema may instead be cast as a subphonemic-level category of VOT realization.

situational context-level schema, episodes from the *parrot* category of the word-level pronunciation schema, episodes from the /p/ category of the phoneme-level schema, and episodes from the voiceless stop category of the class-level schema. As all four of these categories now have a longer mean VOT than in the beginning of the experiment, the resulting production of "parrot" will most likely have a longer VOT than a production from the beginning of the experiment would have. The word-level representation schema will only confirm the identity of the word being produced.

If the word to be produced is not one that was heard during the experiment, the matching category in the word-level pronunciation schema will not have a longer mean VOT than it did at the beginning of the experiment. So if the word to be produced is "puffin", the participant will draw on episodes from the PUFFIN category of the word-level pronunciation schema, although the same episodes from the other three schemata will be activated as for "parrot". Since the word-level schema is only one of four relevant schemata, this difference will not figure heavily in the resulting production of "puffin", which will still most likely have a longer VOT than a production from the beginning of the experiment would have. And if the word to be produced began with a different phoneme from the manipulated phoneme, only two of the pertinent schemata would contribute a longer mean VOT than before the task. As a result, "kazoo" would have a longer VOT because of the adjustment to the voiceless stop category only, but would not draw on the longer mean VOT of the /p/ phoneme-level category, meaning that the difference in length would be smaller than for a word beginning with /p/.

In the other condition of this experiment, in which participants were exposed to a model talker with shorter-than-average VOT, the same five categorization schemata come into play, with word-level representations being sidelined as before. However, the episodes of /p/ with shortened VOT are not stored in the phoneme-level category for /p/, as they fall outside of the definition of the category. Likewise, they are not stored in the natural class-level category for voiceless stops, as they fall outside of the definition of that category. As a result, mean VOT will not change from the beginning of the experiment, and no change to mean VOT will be evinced in post-task production.

1.4 This dissertation

This first chapter constitutes the outline of a theory that treats convergence across disparate domains of linguistic structure as an empirical pattern that falls out naturally from the assumption of episodic recall and storage, and the assumption that multiple representations of the same input are stored separately and recalled independently. Episodic storage is constrained by the categorization schemata that are currently active; each episode is stored in all active categories whose definitions it is perceived to match. In this approach, convergence is not a discrete process, and it may indeed not be confined to language. Dias & Rosenblum (2011) found greater convergence when conversation partners could see each other than when they could only hear each other; Gentilucci & Bernardis (2007) found that speakers converge in lip movement as well as acoustic features when they can see their interlocutor. These findings point to the domain-generality of convergence.

In the following chapters I pursue the categorization schema account. Chapter 2 focuses on the assertion that multiple representations of the same episode can be stored under different categorization schemata. In it, I report the results of a study that looks at the effects of lexical

categories on phonetic convergence by comparing convergence to words with convergence to phonetically similar pseudowords.

Chapter 3 describes an experiment designed to test a pair of predictions of this theory related to the influence of social factors on the activation of categories. The first prediction is that social factors that affect convergence will do so in similar ways on phonetic, lexical, and syntactic levels of linguistic structure, so a given speaker should show a similar pattern of convergence across these levels, assuming multiple levels are active at once. The second prediction is that personal empowerment correlates inversely with social inhibition, such that speakers reporting higher levels of empowerment will converge more than speakers reporting lower levels of empowerment.

In Chapter 4, I discuss the ramifications of these findings for theories of sound change, and report the results of an experiment illustrating that accommodation can directly result in the appearance of new variants within an interaction. Chapter 5 comprises a summary and conclusion.

Chapter 2: Episodic storage and recall

2.0 Introduction

This chapter focuses on the assertion that multiple representations of the same episode can be stored under different categorization schemata. Here I report the results of a study that looks at the effects of lexical categories on phonetic convergence by comparing convergence to words with convergence to phonetically similar pseudowords.

One of the starting assumptions of the proposal in this dissertation is that convergence to *the way something is being done* is a consequence of storing and recalling experiences in episodic storage. As new exemplars are stored in the same category that defines the production of one's own exemplars, a new exemplar has a greater effect on a category with smaller membership than on a category with more members. Because of this, one of the ramifications of episodic memory is the effect of frequency on rate of convergence. In the previous chapter, I discussed[14] studies pointing to the impact on convergence rate of absolute and relative word frequency (Goldinger 1998; Goldinger & Azuma 2004; Nielsen 2011, 2014), as well as the effects of time passing within an experiment (Pardo 2006). These findings all show greater rates of convergence to lower-frequency words than to higher-frequency words.

Within the context of this assumption, there is an outstanding question of when and how new lexical exemplar categories are created. Vaan et al. (2007) investigated the processing of neologisms using a productive suffix in Dutch. They found an effect of lexical priming on morphologically complex derived forms after a single exposure, but were agnostic as to whether this was the result of a whole-word lexical representation, or a trace of the particular combination of morphemes. In the event that someone hears a word for the first time, do they automatically create a new category for it? A neurolinguistic study by Fiebach et al. (2002) found similar brain activation between low-frequency words and pseudowords compared to high-frequency words, suggesting that low-frequency words are processed in similar ways to pseudowords. If this means that new word-level categories are immediately created, we would expect a very high rate of convergence to novel words: the speech being converged to is the new lexical category whose membership constitutes exactly the speech that the converger just heard. If a new lexical category is not immediately created, we would expect no convergence due to the effect of lexical exemplars.

However, an additional ramification of our starting assumption (that convergence is a consequence of episodic storage and recall) is that empirical differences in convergence behavior may be due to differences in the simultaneous recruitment of information from exemplar categories at multiple levels of linguistic representation. Various studies have previously shown results indicating such multiple recruitment (Nye & Fowler 2003; Nielsen 2011, 2014). In this chapter, I survey previous findings that suggest a complex relationship between categories at different levels of representation, and interpret these findings within the current approach. I then report the results of a study targeting phonetic convergence in a task in which listeners shadowed high-frequency, low-frequency, and nonce words in a Mechanical Turk task.

[14] See Section 1.1.1.

2.1 Interacting levels of convergence

Previous studies have looked at convergence at many levels of linguistic structure, ranging from a single phonologically specific phonetic context (Kraljic, Brennan, & Samuel 2008) to discourse-level description schemes (Garrod & Anderson 1987). Given the assumption that convergence is a consequence of episodic storage and recall, empirical differences in convergence across individuals or contexts may be due to differences in the simultaneous recruitment of information from categories across multiple active schemata. This section constitutes a cursory review of some studies that support this idea.

Kraljic, Brennan, & Samuel (2008) looked at accommodation to a backed /s/ phoneme in two experimental conditions. In one condition, ambiguous [sʃ] phones occurred in all contexts of an individual's speech. In the other condition, [sʃ] only occurred in /str/ clusters, as is the case in some dialects of American English. Participants in the latter condition showed convergence in the spectral mean of their /s/ productions in the /str/ context, whereas subjects in the idiolectal-/s/ condition did not. The authors interpreted this result to mean that exposure to the idiolectal [sʃ] does not lead to convergence to that variant, and they claimed that this finding is evidence against an inherent link between production and perception.

However, there are two different reasons why a categorization schema account predicts the results that Kraljic et al. found. First, the distribution of /str/ is much more restricted and less common in English than the distribution of /s/. As such, the crucial [sʃ] phones compose a greater proportion of episodes in the /str/ category than in either candidate category (/s/ or /ʃ/) of the relevant schema. As such, a categorization schema account predicts drastically less adjustment to /s/ across all contexts when the phoneme-level schema is active, than to /s/ specifically before /tr/ when the phonetic environment-level schema is active. Second, the subjects were all from an area where the /s/ backing in an /str/ context was a common feature, and many of the subjects had this feature themselves, although they were not necessarily aware of it. A dialect-specific categorization schema would therefore already include backed-/str/ information, meaning that convergence had an extant target. On the other hand, /s/ backing in all contexts is not a common feature in the subjects' experience, meaning that no dialect-specific category was relevant, and this information did not inform subjects' own production of /s/ from within the dialect-specific schema.

Nielsen (2011) exposed participants to model speech with extended voice onset time (VOT) exclusively in words beginning with /p/, and then recorded them saying words beginning with both /p/ and /k/. Participants extended their VOT in words beginning with both phonemes, although new words with initial /p/ evinced a higher degree of VOT adjustment than words with initial /k/. As discussed in Chapter 1 (Section 1.8), this finding can be interpreted as indicating the existence of three separate categorization schemata, at the word, phone, and phonetic feature levels.

Brennan & Clark (1996) examined lexical entrainment in dyads — pairs of participants were asked to convey the layout of a set of cards depicting objects that either required individuation (i.e. different types of shoes, different breeds of dogs) or did not (one shoe with one dog and one other thing). The authors found that speakers incrementally converged on names for these objects, which names they called conceptual pacts. They concluded that a *historical* account of lexical entrainment was necessary, in which recency, frequency, and partner specificity entered into

a speaker's decision to use a particular variant. Under the current approach, the results of this study provide support for a dynamically updated interlocutor-level categorization schema alongside a referent label-level schema.

If multiple categorization schemata are active at once and can independently store the same episodes, we should be able to test the intervention of a particular schema by setting up a situation in which that schema lacks an appropriate category for some episodes. We can do this by testing convergence at the lexical level, with words that exist and presumably have extant categories, and with words that do not heretofore exist and presumably do not have extant categories. Either new categories will be created immediately upon exposure to the new lexical item, or they will not. If new categories are formed, the only episodes within those categories will be those just heard for the first time, so we would expect a relatively strong convergence effect toward those episodes. Conversely, if new categories are not formed, we should expect less convergence toward pseudowords, as there is no reinforcement of convergent material from the lexical categorization schema. In one of a very small number of studies that examined imitation in a phonetic feature and did not find it, Mitterer & Ernestus (2008) showed that Dutch speakers imitated the presence of pre-voicing in producing initial voiced stops in pseudowords, but did not imitate the amount of pre-voicing. This non-convergence may be related to the lack of lexical-level categories.

Goldinger (1998) conducted a series of phonetic imitation experiments investigating the differences between immediate shadowing and delayed shadowing. Using an AXB perceptual similarity task, Goldinger found higher perceived similarity in immediate shadowing than in delayed shadowing, as well as higher similarity after more repetitions; these findings held true for high-frequency, low-frequency, and pseudowords. However, for both production and perception tasks, pseudowords and actual words were examined in different experiments and never co-occurred, meaning that participants in each task knew to expect only words or only pseudowords. This approach does not suit for our purposes for two reasons. First, we cannot directly compare patterns of convergence within individuals for words versus pseudowords. Second, given the differing expectations of speakers in word versus pseudoword experiments, it is possible that participants were weighting different levels of linguistic representation differently across the two types of tasks. So if participants in a shadowing experiment encountered both words and pseudowords in the same task, they might use lexical representations less than they would in a task with only words.

2.2 Episodic storage and social categories

Previous studies have found effects of social factors on convergence rates. Several previous studies (Namy et al. 2002, Pardo 2006, Pardo et al. 2010, Aguilar 2011, Babel 2012, Walters et al. 2013) have shown effects of speaker sex on the degree of phonetic and phonological convergence. In each of the listed studies with a task involving recorded speech (Namy et al. 2002, Babel 2012, Walters et al. 2013), female participants converged more than male participants; in the studies with a task involving natural dyadic conversation (Pardo 2006, Pardo et al. 2010, Aguilar 2011), male participants converged more than female participants. In other words, men converge more when there is a social context, whereas women converge more when there is not.

One possible explanation for this pattern is that men have fewer episodes stored in their social context categories, such that new socially-categorized episodes hold greater sway.

Additionally, all of the mentioned studies looking at social factors were carried out in Western social contexts, in which men stereotypically hold positions of greater power than women. Elevated power is associated with increased attention to stereotypes (Goodwin et al. 2000), while reduced power is associated with the use of individuating information, such as particular social contexts (Fiske 1993). Assuming that recalling a stereotype is incompatible with accessing individuating information from experience, more powerful individuals may bypass their context-level categories more often in accessing stereotypes. As such, women taking part in convergence studies with a social component may converge less because they are using a denser category than men are.

2.3 Predictions

Our first prediction (H1) is that speakers will show greater convergence toward low-frequency words than high-frequency words, as previous studies have found (Goldinger 1998; Goldinger & Azuma 2004; Nielsen 2011, 2014).

Our next set of predictions regards speaker's convergence toward nonce words. If new lexical categories are immediately created upon hearing a new word, we would expect greater convergence toward nonce words than either low-frequency or high-frequency words (H2). However, if nonce words are instead received and processed as purely phonetic/phonological material, we would expect less convergence toward nonce words than high- or low-frequency words (H3).

Assuming that exposure to pseudowords does not immediately spur the creation of new word-level categories, there are two hypotheses regarding the underlying cause of convergence in high-frequency words which can be tested. If high-frequency words exhibit convergence *because* there is a meaningful shift in word-level categories, we should expect high-frequency words to show larger convergence effects compared to pseudowords, which have no word-level categories (H4). On the other hand, if high-frequency words exhibit convergence *despite* the lack of a meaningful shift in word-level categories, we should expect high-frequency words to show smaller convergence effects than pseudowords, which have no inertia from word-level categories (H5).

Our final set of predictions regards the effects of social categories. If socially-defined exemplar categories are always active, we would expect differences between speakers along social measures such as personal power to have an effect on convergence rates (H6). However, if social categories require a social context to be active, we would expect no such differences (H7).

2.4 Experiment

2.4.1 Methodology

Participants completed a Human Intelligence Task (HIT) on Amazon Mechanical Turk in which they requested qualification for the speech recording task. In order to receive the necessary qualifications, participants had to report their age as between 18-35; that they had an external microphone and headphones; that they had no hearing loss; that they were native speakers of English; and that they consented to have their voices recorded. Participants who self-reported all of these qualifications were enabled to complete a second HIT which was the experiment.

In the second HIT, participants (n = 45) filled out a survey regarding their personality traits and sense of personal power in relationships with others (from Anderson et al. 2012). Participants then navigated in their web browser to a webpage with an interface allowing them to start and stop recordings of their voice. The webpage began with a page of text instructions for using the interface. When participants clicked the "start recording" button, a word appeared on the screen for them to say; when they subsequently pressed "stop recording", the word disappeared. Participants proceeded at their own pace through 80 words in this manner, at which point another page of text instructions appeared for the next section. For the second section, instead of words appearing when "start recording" was pressed, participants heard an audio recording of a voice and repeated the word they heard. This section included 120 words in a random order that was fixed across participants: 40 high frequency words, 40 low frequency words, and 40 nonce words[15]. All of the words and pseudowords were disyllabic with primary stress on the initial syllable. Word frequencies were obtained from the SUBTLEX corpus (Brysbaert & New 2009). Participants were informed that some of the words would be unfamiliar: "Some of the words you hear are not real English words — they are words from children's books, etc. Please repeat them as you hear them." After all 120 words, another page of text instructions appeared for the next section. The third section was the same as the first except that the words presented were in a different order.

The model talker for this section was a male native speaker of Southern California English. The model talker produced the entire list of 120 words three times through, and the third reading was used as stimuli for the shadowing task, in order to normalize phonetic reduction on all three categories of words.

Five participants' data had to be excluded due to issues with the recordings: two participants consistently missed the recording window, two had choppy or empty recordings, and one was inaudible. The remaining (n = 40) participants' recordings were aligned using a modified version of the Penn Forced Aligner (Yuan and Liberman 2008). Measurements of the first two formant frequencies were taken at the temporal midpoint of each vowel and normalized using the Bark scale.

In order to quantify convergence behavior, I calculated the Euclidean distance in F1/F2 vowel space of participants' vowel productions from the mean formant values of the model speaker's productions of that vowel that participants heard in the shadowing condition. I then measured the difference between these Euclidean distance values for the pre-exposure block and those for the post-exposure block, and took the difference in distance between the two blocks as the outcome variable, following the method described in Babel (2012). I created a linear mixed-effects regression model using the lme4 package (Bates et al. 2016) in R (R Core Team 2016) with difference in distance as the outcome variable; fixed main effects of vowel (among AA AE IH IY OW UW, reference level AA), frequency class (high-frequency, mid-frequency, and low-frequency), reported power (among high, medium, and low self-ratings), and sex; an interaction effect of frequency class and reported power; and random intercepts for participant and word. The interaction between frequency class and power, if significant, would capture the possibility that individuals' power influences the relative density of their stored episodes for high-frequency versus low-frequency words.

[15] See Appendix A for a list of the words used and their frequencies, and Appendix B for the nonce words used.

Because each pseudoword was only produced once during the shadowing task, and not repeated at any other time during the experiment, the difference in distance measure could not be used in comparing convergence behavior between words and pseudowords on a by-word basis. As such, I used the absolute Euclidean distance between participants' vowels and the model's vowels in the shadowing condition as a stand-in measure, in order to observe whether participants' pseudowords patterned more like high-frequency words or low-frequency words. I created a second linear mixed-effects regression model with Euclidean distance from the model's F1 and F2 for the given word as the outcome variable; fixed main effects of vowel (among AA AE IH IY OW UW, reference level AA), frequency class (high-frequency, low-frequency, and nonce), reported power, and sex; interaction effects of frequency class and vowel, and frequency class and reported power; and by-participant random slopes for frequency class and vowel. An interaction between frequency class and vowel, if significant, would capture differential vowel reduction effects based on word frequency; an interaction between frequency class and power would capture the possibility that individuals' power influences the relative density of their stored episodes for high-frequency versus low-frequency words. A model including random slopes for vowel failed to converge, so these were not included.

2.4.2 Results

For the analysis of difference in distance between vowels in the pre-exposure task and vowels in the post-exposure task, no included fixed effect was a significant predictor of difference in distance. After successively removing the least informative predictor according to the Akaike information criterion, the resulting model included no fixed effects, and only a random effect of participant.

For the analysis of Euclidean distance between participant and model vowels in the shadowing task, the inclusion of sex, reported power, and the interaction between the vowel and frequency class, as well as by-subject slopes for frequency class, each did not improve the model, so were removed during model comparison. The coefficients of the resulting model, of Euclidean distance in the shadowing task as a function of word frequency class and vowel with random intercepts for subjects, are shown in Table 2.1; *p*-values were supplied by the lmerTest package (Kuznetsova et al. 2016). Lower numbers indicate smaller distances between the model talker's vowels and participants' vowels.

Overall, participants' vowel productions were an average of 2.2 on the Bark scale from those of the model talker. Participants' /æ/ vowels were significantly closer to the model talker than other vowels by about 1 (β = –0.99, SE = 0.08, $p < 0.001$), whereas /i/, /oʊ/, and /u/ were farther away by about 1.15, 0.5, and 0.79, respectively (/i/: β = 1.15, SE = 0.08, $p < 0.001$; /oʊ/: β = 0.47, SE = 0.08, $p < 0.001$; /u/: β = 0.79, SE = 0.08, $p < 0.001$). Participants' vowels were closer to those of the model speaker in low frequency words and nonce words than in high frequency words; however, this effect was not statistically significant.

Table 2.1: Vowel Euclidean distance from model talker in shadowing task

```
Random effects:
 Groups   Name        Variance Std.Dev.
 Word     (Intercept) 0.04674  0.2162
 Subject  (Intercept) 0.04898  0.2213
 Residual             0.67851  0.8237
Number of obs: 4463, groups:  Word, 118; Subject, 40

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)   2.20235    0.07245 154.14000  30.398  < 2e-16 ***
FclassLow    -0.06827    0.05768 117.22000  -1.184    0.239
FclassNonce  -0.03696    0.05728 116.81000  -0.645    0.520
VowelAE      -0.99394    0.07750 123.13000 -12.824  < 2e-16 ***
VowelIH       0.03615    0.07616 125.39000   0.475    0.636
VowelIY       1.15246    0.07821 121.23000  14.735  < 2e-16 ***
VowelOW       0.47063    0.07817 137.41000   6.021 1.50e-08 ***
VowelUW       0.79293    0.08475 120.17000   9.356 4.44e-16 ***

Correlation of Fixed Effects:
            (Intr) FclssL FclssN VowlAE VowlIH VowlIY VowlOW
FclassLow   -0.392
FclassNonce -0.415  0.503
VowelAE     -0.518 -0.005  0.019
VowelIH     -0.533  0.012  0.034  0.487
VowelIY     -0.502 -0.025 -0.002  0.476  0.482
VowelOW     -0.495 -0.012  0.015  0.462  0.468  0.458
VowelUW     -0.473 -0.004  0.015  0.440  0.445  0.436  0.423
```

2.4.3 Discussion

Hypotheses.

H1: Results were in general accordance with previous studies looking at the effects of lexical frequency on phonetic convergence, in that participants' vowels in low-frequency words were more similar to a model talker's vowels compared to high-frequency words. However, this pattern is also in accordance with previous findings regarding frequency effects on vowel space reduction (Aylett & Turk 2006), and the current study cannot disambiguate between these two interpretations of the observed pattern.

H2-H5: Nonce words did not fall into any clear pattern in relation to either high-frequency or low-frequency words. As such, none of the hypotheses pertaining to nonce words were borne out. It is possible that differences in the model talker's pronunciation between the three categories of words had an effect. However, it should be noted that the nonce words were presented as real lexical items without a relevant referent. While participants were told that the unfamiliar words came from children's books, there were no examples given and no contextual reason for participants to assign any meaning to the nonce words they heard. If the definition of a lexical

category requires a semantic component, participants would not be able to create new lexical categories for nonce words under these circumstances.

Table 2.2: Hypotheses under consideration for frequency-convergence interaction

H1: Speakers will show greater convergence toward LF words than HF words.
H2: Speakers will show greater convergence toward nonce words than LF or HF words.
H3: Speakers will show less convergence toward nonce words than LF words.
H4: Speakers will show less convergence toward nonce words than HF words.
H5: Speakers will show greater convergence toward nonce words than HF words.
H6: Personal power will affect convergence rates.
H7: Personal power will not affect convergence rates.

H6/H7: There was no effect of participant self-rating of personal power, in support of H7 which predicted no effect of social factors due to the asocial nature of the task. While it is possible that the model talker's accent may have introduced a social dimension anyhow, particularly if participants processed more and less accent-specific vowels differentially, no such effect was observed.

2.5 Conclusion

The nature of a study looking at nonce words is such that the available data is sparse for any given word. The second repetition of a given pseudoword is intrinsically different from the first. While the results of this study did not contradict previous findings that low-frequency words show greater convergence effects than high-frequency words, they also did not confirm those findings. There may not have been sufficient statistical power to contribute anything further about the specific questions at issue.

Chapter 3: Social effects on convergence

3.0 Introduction

There are two ways that social factors can affect convergence. First, they can affect perception of episodes directly (Strand and Johnson 1996, Strand 1999, Niedzielski 1999, Sumner and Samuel 2009). Second, they can affect which categorization schemata are active. They can do this on either a long-term scale by determining the makeup of an individual's categorization schemata in the first place, or on a short-term scale by determining which schemata are activated within a given socially-informed context. The current chapter discusses the effects of social factors on the activation of categorization schemata within the context of a conversational experiment.

Although many if not all levels of linguistic structure evince convergence, there are differences between the ways that categorization schemata are organized. For instance, speakers must choose between phonological variants within a given category when speaking, but they may sidestep all competing lexical or syntactic variants in favor of talking about something else. If a speaker has a reason to keep a referent's gender hidden in English, for example, they can avoid using pronouns at all, e.g. by choosing instead to use passives. Likewise, if a particular level of semantic representation is inappropriate for a referent, speakers can use a different level (e.g. *person* for *man* or *woman*). However, the same option is not easily available at the phonological level. While speakers may elect to avoid a particular variant (e.g. a final glottal stop for /t/ in English), it is much harder to avoid words with final /t/ altogether in order to escape making such a determination.

There is an additional potential asymmetry in convergence, in that it is not necessarily the case that different individuals will necessarily evince the same type of accommodative behavior in the same situation. Differences in accommodation may stem from personal factors including but not limited to: prior experience with the linguistic features being encountered; intensity of desire to establish social acceptance or identity, in the context of the given interlocution as well as more generally; physiological or neurological idiosyncrasies directly impinging on speech perception and/or production; fatigue; inattention; and emotional state. Some of these factors are potentially useful in predicting accommodative behavior. Researchers have investigated the effects of factors such as social identification (Bourhis & Giles 1977), liking (Babel 2012), power (Pardo et al. 2010), and empathy (Abrego-Collier et al. 2011) on whether or not, and to what extent, speakers display accommodation effects.

There is then a twofold question, of whether social factors that affect convergence will do so in similar ways across different levels of linguistic structure, and whether they will do so in similar ways from one individual to the next. This chapter describes an experiment designed to explore both parts of this question, investigating the influence of social factors on the activation of categories. First, given the differences between categorization schemata, both in terms of organization and in terms of applicability to any particular situation, should we expect convergence to ramify in the same way at every level of linguistic structure? Second, are there specific social factors that affect convergence in specific ways, independent of an individual's history and experience with language? The experiment discussed in this chapter explores individuals' sense of personal empowerment as one such candidate factor.

3.1 Differential category activation

Pickering and Garrod (2004) proposed Interactive Alignment Theory (IAT), a mechanistic account of dialogue in which alignment between speakers is reflexive and consequent. Under their account, alignment results from "implicit common ground", which is built up between interlocutors over the course of a dialogue. Individuals in conversation align their situation models in such a manner that their production and comprehension systems interact at all levels of linguistic processing. This alignment is driven by priming, in that every utterance automatically primes all associated representations for both speakers. Their six starting points are (formatting mine):

"(1) Alignment of situation models [...] forms the basis of successful dialogue;
(2) The way that alignment of situation models is achieved is by a primitive and resource-free priming mechanism;
(3) The same priming mechanism produces alignment at other levels of representation, such as the lexical and syntactic;
(4) Interconnections between the levels mean that alignment at one level leads to alignment at other levels;
(5) Another primitive mechanism allows interlocutors to repair misaligned representations interactively; and
(6) More sophisticated and potentially costly strategies [...] are only required when the primitive mechanisms fail to produce alignment."

Points (2) – (4) are the relevant ones to the current discussion, as the categorization schema account pursued here focuses on component episodes of a dialogue rather its overall success. This account disagrees with (2): alignment is not achieved by a specific priming mechanism; it is a consequence of episodic storage and recall. Mechanism aside, it agrees with (3) in that different levels of representation evince alignment for the same fundamental reason. Point (4) is what is currently at issue.

In support of (4), Pickering and Garrod point to the *lexical boost* effect. This refers to a lexically-driven increase in the rate of syntactic persistence. Hartsuiker et al. (2008) define the lexical boost: "If the target sentence uses the same verb as the prime sentence, there should not only be priming because of the combinatorial nodes' residual activation, but also because of the extra activation traveling from verb to combinatorial node via the active link. Thus, there should be more priming when prime and target have the same verbs than different verbs" (2008:215-16). Branigan et al. (2000) looked at syntactic priming in the dative alternation in English, and found that speakers were more likely to reuse a syntactic structure when the new use shared a verb with the priming structure. However, in looking at the cognate structure in Dutch, Hartsuiker et al. (2008) found that syntactic priming outlasts the lexical boost effect: with a longer lag between the prime and the target utterance, speakers no longer show a lexical boost, although they still use the primed structure at an increased rate. Hartsuiker et al. interpreted this finding to indicate that a combination of short-term and long-term mechanisms are at play.

There is a difference here though, in that conversational data do not involve lexical targets *per se*. Previous studies investigating lexical effects on syntactic persistence have specified via one

mechanism or another which word speakers should use; but in natural conversation, speakers are free to use any verb they want. The lexical boost hypothesis has no opinion on whether increased syntactic priming is dependent on the verb actually being produced, or whether it is dependent on the verb being applicable to the context at hand. In the event that words and syntactic structures are being recalled in discrete processes, there may not be a lexical boost.

The current approach treats syntactic-level and word-level convergence phenomena as being of the same kind, but manifesting differently due to their different circumstances. Since syntactic structures (in English) are generally made up of multiple words, words are necessarily more frequent in general than syntactic structures. As such, within any given context there are more word-level episodes than syntactic phrase-level episodes, so any given word-level episode will be superseded by new word-level information more quickly than a phrase-level episode will be. Unlike IAT, however, there is no necessary direct link between the word- and phrase-level categorization schemata; the fact that both are populated by episodes from the same stretch of discourse is what leads to similar behavior between them.

Also unlike IAT, the current account does not require that interlocutors construct or maintain situation models. Such an occurrence may well fall out as a result of everything else lining up, but it is a consequence of the automatic accommodation process, rather than its cause. In most circumstances, all interlocutors within a conversational context may well end up having the equivalent set of active categorization schemata active, resulting in cross-level alignment akin to that proposed by IAT.

It is also possible, however, that social, cultural, or attentional differences between interlocutors may lead to inequivalences between their active categorization schemata[16]. I have already discussed the potential effects of attention[17] and experience[18] on convergence. If social dynamics lead to different interlocutors paying attention to their interaction in different ways, there may be a mismatch in active categorization schemata leading to different rates of convergence between them within a given category.

The experiment detailed in this chapter is in part an investigation of Pickering and Garrod's (2004) point (4) above. Specifically, I am looking at whether alignment at phonetic, lexical, and syntactic levels of representation occurs at a relative rate for a given individual within a given interaction.

Some general trends have been found that suggest that we should expect similarities across levels. Using the English dative alternation, Rowland et al. (2012) found a larger lexical boost effect in syntactic priming in adults than in children, and in older children than in younger children. However, they found a larger abstract structural priming effect among younger children than older children or adults. This finding is analogous to Nielsen's (2014) finding regarding phonetic convergence toward lengthened VOT (discussed in Chapter 1.4). Nielsen (2014) looked at VOT accommodation in preschoolers, third graders, and college-age adults, and found that both groups of children extended VOT by twice as much as adults did. Furthermore, whereas preschoolers and third-graders showed equivalent VOT extension in /k/-initial words, preschoolers exhibited greater convergence in non-exposure /p/ compared to exposure /p/, whereas third-graders exhibited the opposite pattern. In both Rowland et al. (2012) and Nielsen

16. Or indeed the repertoire of schemata they have.
17. See Section 1.1.3.
18. See Section 1.2.1.

(2014), older people show progressively more specific convergence, whereas younger people show progressively more general convergence. Taken together, the findings of these two studies suggest a parallel between lexical priming and phonetic convergence.

However, unlike phonetic convergence, lexical convergence deals with discrete differences. If two people already use the same word, there's no way to converge. Because of this, convergence can only occur when interlocutors start at different places. As such, it's impossible to tell upon an individual's first reference to a referent whether they were going to have used that label anyway (i.e. they started at the same place), or whether they are accommodating to their interlocutor.

Regarding the systematicity of lexical entrainment, it's almost certainly not the case that all word choices are instances of an automatic process. However, it may be the case that the adoption of palatable terms into a listener's model of the discourse-shared context is automatic. This is all in all a good thing, as there's no reason for such incorporation not to be automatic. Once a term has been examined and found unobjectionable (i.e. there is a one-to-one mapping between the term and a single plausible referent), there's no cognitive reason to go searching for another term.

In examining syntactic persistence as an accommodation effect, certain aspects of previous approaches to both syntactic priming and accommodation need to be reconciled. Syntactic priming is generally conceived of as occurring either within an individual's speech or across individuals in an interaction; accommodation is generally only concerned on the latter case.

In looking at syntactic accommodation, it is difficult to disentangle syntactic changes from lexical changes. Syntactic structures that vary without any lexical differences make up a small proportion of all of the syntactic structures that speakers use. The list of differences between (A) *I would like a cookie* and (B) *Can I have a cookie?* clearly includes syntactic differences, but also clearly includes non-syntactic ones. Table 3.1 lists four or five obvious differences between these two sentences, of which I would characterize (1) as a syntactic difference and (2-3) as lexical differences:

Table 3.1: Differences between (A) *I would like a cookie* and (B) *Can I have a cookie?*

1. A has the auxiliary after the subject; B has the auxiliary before the subject
2. A has the auxiliary *would*; B has the auxiliary *can*
3. A has the verb *like*; B has the verb *have*
4. A is semantically a statement; B is semantically a question
5*. A has level intonation; B has rising intonation

It is also obvious that these four or five differences are not independent of one another. (C) *Can I like a cookie?* is not a likely sentence in any English I know of; (D) *Would I like a cookie?* is a whole other thing pragmatically. This dependent relationship between syntactic and lexical differences makes it difficult to isolate syntactic convergence from lexical convergence. If I use *would* because you used *would*, that will affect which licit syntactic structures I have available to me[19].

Gries (2005) used a corpus-based approach to look at syntactic persistence. He used the ICE-GB corpus of spoken and written British English to look at two constructions, the dative

19. This recalls the *lexical boost effect* (Hartsuiker et al. 2008), which is a robust finding that shared lexical material increases the rate of syntactic persistence.

alternation (*the chorister offered a walnut to the charwoman* vs. *the chorister offered the charwoman a walnut*) and the phrasal verb particle placement alternation (*the chorister put up five dollars* vs. *the chorister put five dollars up*). For each of these two construction types, Gries compared each sequential pair as a prime-target pair; he included pairs of any proximity within a text, and included sequential pairs produced by both the same and different individuals. Gries looked at 3003 prime-target pairs of the dative alternation and 1797 prime-target pairs of the particle placement alternation. Gries extended Bock's (1986) findings of syntactic persistence in the dative alternation to naturalistic data, and found that the particle placement alternation also evinced persistence effects. Further, Gries found a lemma-specific effect: for the dative alternation, particular verbs (e.g., *give, show, offer*) showed priming for the ditransitive form but not the prepositional form; other verbs (e.g., *hand, sell*) showed priming for the prepositional form but not the ditransitive form; and other verbs (e.g., *lend, send*) showed priming for both forms. Similarly, for the particle placement alternation, some verbs (*e.g., take up, find out*) showed priming for only the pre-verbal particle, other verbs (*e.g., put in, take out*) for only the post-verbal particle, and still other verbs (*e.g., pick up, put down*) for either form.

Gries (2005) interprets these results as support for Pickering and Branigan's (1998) findings that priming is stronger when the verb is the same in both prime and target. However, it is possible that this similar pattern has a distinct cause in the two studies. The interaction in the corpus study between verb identity and persistence rate may be due to the bias inherent to each particular lemma, such that, *e.g.*, two instances of the same ditransitive-biased verb are more likely to both be produced in a ditransitive form. However, whereas verbs in the dative construction were (presumably) autonomously generated in Gries's corpus study, the verbs in Pickering and Branigan's study were given to participants as part of the task. In other words, Pickering and Branigan's results are predicated on speakers first selecting a verb and then selecting an appropriate form of the dative construction. If these steps usually occur in the opposite order in natural speech production, or at the same time, the lexical boost effect may not map to natural speech.

These are not quite the same thing. Under the current account, the corpus study has an interaction between identity across successive verbs and priming rate because of the flexibility of particular verbs, or because some verbs are biased toward the more common form of the construction. The fact that Gries (2005) found no main effect of verb identity supports this interpretation. In contrast, the study in Branigan et al. (2000) has an interaction between verb identity and the particular form of the dative.

One difference between Gries's (2005) corpus study and the present study is that the former was concerned with the incidence of syntactic persistence in specific constructions, whereas the current study is concerned with the incidence of syntactic persistence in specific individuals. As we do not know whether all syntactic structures evince persistence effects at the same rates, measuring individual levels of syntactic convergence is a murky issue. My effort here is to select the possible loci of variation that I as an observer find to be most salient, in the hopes that my intuitions here align somewhat with the perceptions of the participants in this study.

3.2 Predictions

There are two main questions that the current experiment is designed to address. First, do speakers show similar patterns of convergence at different levels of linguistic structure within the

same interaction? If different levels of linguistic structure are directly connected as in Interactive Alignment Theory (Pickering and Garrod 2004), we would expect similar patterns of convergence at phonetic, lexical, and syntactic levels of structure, such that a given individual will converge along all three levels at a similar normalized rate within an interaction (H1). However, if levels of linguistic structure are independently informed and convergence is a consequence of separate processes at each level, we would expect differences from one level to the next contingent on each level's particular situation within the interaction (H2). Another possibility is that there is a two-way distinction between convergence that uses procedural memory and convergence that uses episodic memory, such that phonetic and lexical convergence would show a similar pattern that may differ from syntactic convergence for a given individual (H3).

Second, is convergence affected by an individual's sense of personal empowerment? According to Keltner et al., elevated power is associated with positive affect, which increases the likelihood of automatic (which I read as *reflexive*) social cognition (2003:272). Power is also associated with increased attention to stereotypes (Goodwin et al. 2000). Categorization schemata are similar to stereotypes in both their purpose of facilitating comprehension of a situation, and in their non-activation in the fact of working memory load. Conversely, reduced power is associated with the use of controlled social cognition and recruitment of individuating information (Fiske 1993). Categorization schemata may be the mechanism through which this individuating information is recruited. If stereotypes and categorization schemata compete for working memory capacity, it is possible that people with elevated power use stereotypes in preference, and so will not have the free working memory capacity with which to recruit the optimal set of categorization schemata for use within a given context.

The question then is whether optimal categorization schemata induce or inhibit convergence. If increased convergence is due to a closer match between input and active categorization schemata, we might expect higher-power individuals to converge less (H4), as they are recruiting stereotypes instead of activating the optimal schemata for the situation. However, if convergence is due to there being less of a match between input and active schemata, such that the active schemata undergo a greater shift, we would expect higher-power individuals to converge more (H5), as they would be less likely to have appropriate schemata active for the given context.

3.3 Experiment

3.3.1 Methodology

29 native English speakers participated in the experiment at UC Berkeley, for US$5. The data from two participants were discarded due to equipment issues seriously affecting the quality of the recordings, leaving 27 participants whose data was analyzed (9 men, 18 women).

The experiment consisted of three blocks: (1) baseline recording, (2) two iterations of a trading game, and (3) a set of background questionnaires. Each session typically lasted between fifteen and thirty minutes.

Participants were brought into a room containing a couch and a chair separated by a low coffee table. They were seated in either the couch or the chair, and given a consent form. The confederate was seated in the other location, and also given a consent form. After signing the

consent form, both the participant and the confederate were given lapel microphones and instructed to read a word list, starting with the participant[20]. At this point the participant had not yet heard the confederate speak. They were instructed to pause for about half a second between each word, in order to minimize the introduction of utterance-level intonational contours. After both people had read the word list, they were given instructions for the game task and the first grids were revealed. Neither person could see the requirements for the other, but they could see the blocks that the other person had.

For each iteration of the trading game, the two players each had a grid of nine requirements, and a set of colored blocks each bearing a picture of an object (from Snodgrass & Vanderwart 1980; see Figure 3.1). Each requirement dictated either the color of the block to be placed in that spot on the grid, or a semantic category for the object depicted on the block. The requirements were such that, although many objects fit many grid spaces, there were relatively few solutions that will satisfy all requirements of both grids at the same time. Moreover, some objects had to be in specific spaces in one or the other grid; because of this stricture, the confederate was able to manipulate the length of the task by either taking those objects or putting them in the wrong space. The participant and confederate were given written instructions to take turns asking for an object until both players simultaneously satisfied all of their requirements. They were specifically informed that this was a cooperative task. The participant's instructions indicated that the confederate would go first.

Figure 3.1: Pictures of a bee/wasp and pan/saucer (from Snodgrass & Vanderwart 1980)

[20] See Appendix D.

Table 3.2: Game experiment objects

Referent	Expected labels	Additional labels
AXE	axe, hatchet	
BEE	bee, wasp	insect
BIKE	bicycle, bike	
CRICKET	cricket, grasshopper, locust	insect
CUP	cup, glass	milk, glass of water
ENVELOPE	/ɑ/nvelope, /ɛ/nvelope	letter, email
GUN	gun, pistol	revolver
HEN	chicken, hen	duck, rooster
PAN	pan, frying pan, saucepan	
PANTS	pants, (trousers)	clothes, pair of pants
PLANE	airplane, plane	aeroplane
PRAM	baby carriage, buggy, carriage, stroller	baby car, pram
RABBIT	bunny, rabbit	bunny rabbit, squirrel
RAT	mouse, rat	rodent
SOFA	couch, sofa	bed
TIE	necktie, tie	
TRASH	garbage can, trash bin, trash can	garbage, trash
TV	television, TV	microwave
VASE	/vɛɪs/, /vɑz/	

Table 3.3: Percentage name agreement (from Snodgrass & Vanderwart 1980)

Concept	% Name agreement
Airplane	60%
Axe	90%
Baby carriage	52%
Bee	60%
Bicycle	88%
Chicken	67%
Couch	67%
Envelope	98%
Frying pan	60%
Garbage can	88%
Glass	98%
Grasshopper	71%
Gun	74%
Mouse	79%
Pants	88%
Rabbit	100%
Television	52%
Tie	69%
Vase	95%

Figure 3.2: The game experiment in progress

Table 3.4: Starting distribution of objects

Subject's start:	Confederate's start:
Plane	Garbage
Buggy	Hen
Wasp	Gun
Cricket	Sofa
Axe	Rat
Vase	Rabbit
TV	Glass
Pan	Tie
Bike	Pants
Envelope	

Table 3.5: Requirements by round/player, Iteration 1

Grid 1:		
Category	**Plausible category members**	**Solution components**
PURPLE	Gun, Envelope, Cricket, Pan	Envelope, Pan
TRANSPORT	Buggy, Bike, Plane	Bike, Plane
GREEN	Hen, Buggy	Hen, Buggy
ANIMAL	Cricket, Bunny, Rat, Wasp, Hen	Bunny, Rat, Wasp
CRICKET	Cricket	Cricket
CLOTHES	Tie, Pants	Tie
WHITE	Glass, Rat, Pants, Sofa, Bike	Pants, Sofa
WEAPON	Axe, Gun	Gun
BLUE	Bunny, Plane, Vase, Axe	Vase, Axe
Grid 2:		
Category	**Plausible category members**	**Solution components**
CONTAINER	Envelope, Pan, Glass, Garbage, Plane, Vase	Glass, Garbage
GREEN	Hen, Buggy	Hen, Buggy
PURPLE	Gun, Envelope, Cricket, Pan	Envelope, Pan
WHITE	Glass, Rat, Pants, Sofa, Bike	Pants, Sofa
TRANSPORT	Buggy, Bike, Plane	Bike, Plane
ANIMAL	Cricket, Bunny, Rat, Wasp, Hen	Bunny, Rat, Wasp
ANIMAL	Cricket, Bunny, Rat, Wasp, Hen	Bunny, Rat, Wasp
BLUE	Bunny, Plane, Vase, Axe	Vase, Axe
CONTAINER	Envelope, Pan, Glass, Garbage, Plane, Vase	Glass, Garbage

Table 3.6: Requirements by round/player, Iteration 2

Grid 1:		
Category	**Plausible category members**	**Solution components**
TRANSPORT	Buggy, Bike, Plane	Bike
ORANGE	TV, Tie	TV, Tie
BLUE	Bunny, Plane, Vase, Axe	Bunny, Axe
GREEN	Hen, Buggy	Hen, Buggy
BLUE	Bunny, Plane, Vase, Axe	Bunny, Axe
KITCHEN	Glass, Pan, TV (Microwave), Garbage	Glass, Pan
STRIPED	Tie, Wasp, Garbage	Wasp, Garbage
CLOTHES	Tie, Pants	Pants
HAS WINGS	Hen, Plane, Wasp, Cricket	Plane, Cricket
Grid 2:		
Category	**Plausible category members**	**Solution components**
PURPLE	Gun, Envelope, Cricket, Pan	Envelope
WEAPON	Axe, Gun	Gun
HAS WINGS	Hen, Plane, Wasp, Cricket	Plane, Cricket
GREEN	Hen, Buggy	Hen, Buggy
KITCHEN	Glass, Pan, TV (Microwave), Garbage	Glass, Pan
LIVING ROOM	Sofa, Vase, TV	Sofa, Vase
LIVING ROOM	Sofa, Vase, TV	Sofa, Vase
YELLOW	Wasp, Garbage	Wasp, Garbage
ORANGE	TV, Tie	TV, Tie

A couple of participants used RAT for the Kitchen block. This enabled many more solutions, as either PAN or CUP became available for PURPLE, with further knock-on repercussions.

After giving the instructions to both players, I moved to the corner of the room where I remained visible to the person on the couch but behind the person on the chair. I spoke mainly to address rules violations (e.g., asking for a white block rather than for a particular white block, taking a block without asking for it, etc.). After either the participant or the confederate indicated that the task was complete, I replaced the grids with a new set of requirements and indicated with a gesture that they should perform the task again. After the second iteration of the task, the participant was given a questionnaire asking for biographical data (age, sex, places of residence, languages spoken, handedness, hearing loss), a Big Five personality questionnaire (Saucier 1994, following Yu 2013), and a personal empowerment questionnaire (Rogers et al. 1997). The confederate was given the same questionnaires and pretended to fill them out[21].

21. One participant remarked that they knew the confederate was not naive because he did not fill out the questionnaires. This participant also stated that they had recently taken part in another study with a confederate, so were anticipating a similar situation here.

The experiment was recorded on a Canon XA10 HD video camera and Sony ECM-77B lapel microphones with Neutrix NC3FXX XLR cables. The video camera was set up to record the table where the game was being played, in order to assist with the reconstruction of turns after the fact in case of ambiguities.

3.3.2 Analysis

To measure convergence for categorical variables such as lexical choice and syntactic structure choice, I used the following paradigm, shown for a lexical variable in Table 3.7:

Table 3.7: Categorical convergence paradigm

Time A — Subject's Form A	bee	bee	bee	bee
Time B — Confederate's Form B	wasp	wasp	wasp	bee
Time C — Subject's Form C	bee	wasp	hornet	wasp
Coded as:	NO	YES	YES	[not coded]

Each variable under investigation was coded in this way, regardless of the number of turns in between Time A and Time B or Time B and Time C. As pilot data indicated that pairs of naive participants without a confederate never changed to a form that had not previously been used, any change between Form A and Form C was coded as convergence as long as Form A and Form B differed. As such, the introduction of a new form was coded as convergent, as it was interpreted as a reaction to the confederate's behavior. If the confederate used the same variable at Time B that the subject did at Time A, the exchange was not coded, as seen in the fourth column.

In order to measure participants' personal sense of power, I used the psychological empowerment questionnaire that participants filled out (from Rogers et al. 1997). This questionnaire separates empowerment into five factors, corresponding to *self-esteem/self-efficacy*, *power-powerlessness, community activism and autonomy, optimism and control over the future,* and *righteous anger*. I normalized each of these factors across all participants ($m = 0$, SD = 1). In addition, I collected power judgments from 18 naive raters using thin-slicing, a technique for drawing objective predictions from short periods of observation of expressive behavior that has been shown to be effective when observing both naturally-occurring and laboratory behavior (Ambady & Rosenthal 1992). I separated out the first 45 seconds of the trading game portion of the experiment for each participant. These short recordings were played for the naive raters who judged the relative power of each participant. The intraclass correlation of the raters' judgments was 90% ($p < 0.001$). I then normalized the judges' ratings as well ($m = 0$, SD = 1).

3.3.2.1 Analysis: Phonetic convergence

I used a modified version of the Penn Forced Aligner (Yuan and Liberman 2008) to align a transcription of each game experiment with the audio portion of the video recordings. I then used a Python script to extract the duration of each stressed vowel produced by the participant during

the game experiment, as well as measurements of the first three formants for each stressed vowel at its temporal midpoint.

3.3.2.1.1 Convergence in stressed vowel duration

I excluded the wordlist reading from the analysis of vowel duration because I expected segment durations to be informed by the difference between elicited speech with pauses specified between words on the one hand, and natural speech within a turn-taking interaction on the other hand. In order to look at the effects of convergence over the course of the experiment, I labeled each stressed vowel produced by the participants as occurring at the beginning, the middle, or the end of the experiment. One participant (S37) produced only 81 stressed vowels over the course of the game experiment; in order to provide a sufficient number of data points for each timepoint, I labeled the first and last thirty stressed vowels as Beginning and End vowels, with all other vowels being labeled Middle vowels. Only two instances of stressed /ɔɪ/ were produced across all participants; they were removed from further analysis. Table 3.8 shows the distribution of vowels across all speakers for the three labeled timepoints.

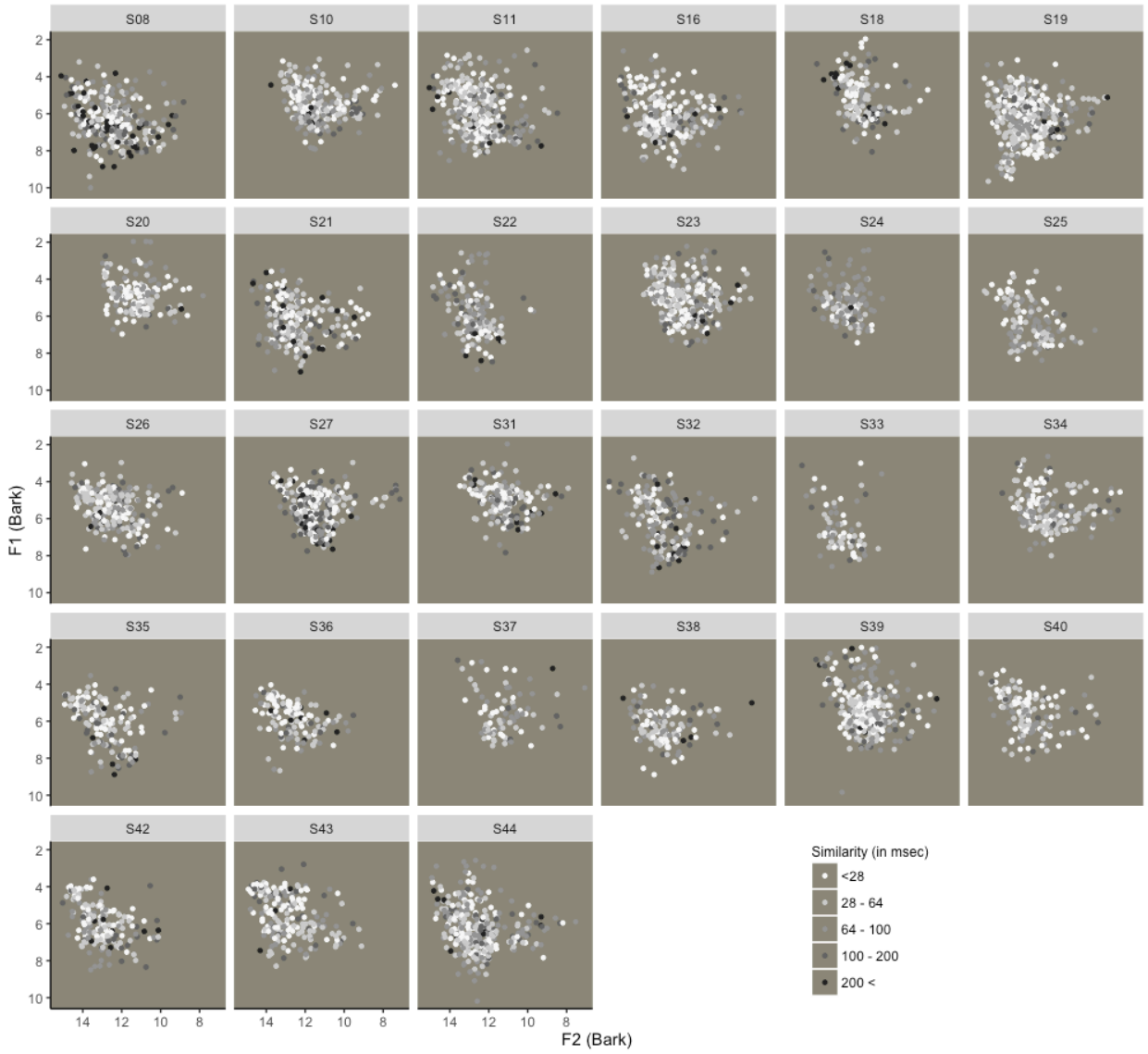
Table 3.8: Distribution of first/last thirty vowels

	[i]	[ɪ]	[eɪ]	[ɛ]	[æ]	[aɪ]	[ə˞]	[aʊ]	[ɑ]	[ʌ]	[ɔ]	[oʊ]	[ʊ]	[u]
First 30	44	45	93	38	149	144	4	5	32	43	12	121	30	49
Middle	211	217	504	304	960	827	29	82	166	143	61	455	262	117
Last 30	42	50	69	66	193	174	6	7	34	31	14	49	53	22

I took the absolute difference in duration between each stressed vowel produced by the participants and the mean duration of the confederate's productions of that vowel within the same span. In other words, I took the difference between each of the participants' first thirty stressed vowels and the confederate's mean duration for that vowel within their first thirty stressed vowels during the game portion of the experiment. Figure 3.3 shows these absolute difference measures for each vowel by subject. In this figure, white dots represent vowels that are within 28 milliseconds (half a standard deviation) in duration from the confederate's mean duration for that vowel.

Unlike the experiment discussed in Chapter 2, Babel's (2009, 2012) difference in distance measure cannot be applied in the context of this analysis because the vowels are not matched between the initial and final phases of the game experiment, and there is no principled way to pair them. By taking the absolute difference in duration, I am allowing for the possibility of patterns of change that may not be clear-cut convergence. For example, if a participant produces a 100-millisecond vowel at the beginning of the experiment and a 140-millisecond vowel at the end, whereas the confederate's mean duration for that vowel is 120 milliseconds, an absolute measure will indicate that no convergence took place.

Figure 3.3: Absolute difference in vowel duration by subject



Difference in duration is shown in the intensity of the points in each participant's vowel space; each token is plotted separately. Darker points indicate vowel productions that were more different from the confederate's mean duration for that vowel.

I created a linear mixed-effects regression model using the lme4 package (Bates et al. 2016) in R (R Core Team 2016), with the absolute difference in duration between the participants' vowel and the confederate's mean for that vowel within the same span as the outcome variable; time (either the beginning, middle, or end of the experiment), sex, vowel identity (with IY as the reference level), and the first and second vowel formants (in Bark) as main effects; and subject and word as random effects. If significant, the effect of time as a predictor would represent changes in the duration of participants', and the confederate's, vowels in relation to one another over the course of the experiment; although such an effect would be ambiguous as far as who was converging to whom, the relative difficulty of the confederate's task suggests that he would have less ability to respond to

participants' vowel duration than the other way around. Participant sex was not a significant predictor and so was discarded from the model. I then added each of the five empowerment factors to this model (SF1: *self-esteem/self-efficacy*, SF2: *power-powerlessness*, SF3: *community activism and autonomy*, SF4: *optimism and control over the future*, and SF5: *righteous anger*), as well as the power judgments provided by naive raters, as a predictor in a series of six models. None of these additions improved the model's Akaike information criterion or were a significant predictor of vowel duration at the $p < 0.05$ threshold; no power factors were therefore added. The resulting model is shown in Table 3.9; *p*-values were supplied by the lmerTest package (Kuznetsova et al. 2016).

Table 3.9: Difference in vowel duration predicted by time point and vowel

```
Random effects:
 Groups   Name        Variance  Std.Dev.
 word     (Intercept) 0.0004425 0.02103
 subject  (Intercept) 0.0001166 0.01080
 Residual             0.0032967 0.05742
Number of obs: 5957, groups:  word, 464; subject, 27

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)  8.728e-02  1.476e-02  1.605e+03   5.914 4.06e-09 ***
F1Bark       1.800e-03  8.039e-04  5.345e+03   2.239  0.02518 *
F2Bark      -2.408e-03  9.261e-04  3.377e+03  -2.600  0.00935 **
TimeMid     -2.037e-03  2.327e-03  5.934e+03  -0.875  0.38134
TimeEnd      2.951e-03  2.975e-03  5.892e+03   0.992  0.32137
vowelIH     -3.498e-02  8.273e-03  2.760e+02  -4.228 3.21e-05 ***
vowelEY     -5.161e-03  8.303e-03  2.270e+02  -0.622  0.53487
vowelEH     -2.006e-02  8.322e-03  2.550e+02  -2.411  0.01662 *
vowelAE     -7.018e-03  8.290e-03  2.020e+02  -0.847  0.39826
vowelAY     -1.050e-02  8.877e-03  2.180e+02  -1.183  0.23802
vowelER     -1.547e-02  1.308e-02  5.390e+02  -1.182  0.23761
vowelAW     -2.333e-03  1.191e-02  2.580e+02  -0.196  0.84485
vowelAA      1.074e-02  9.182e-03  2.990e+02   1.170  0.24306
vowelAH     -5.317e-03  9.135e-03  2.740e+02  -0.582  0.56104
vowelAO      4.761e-03  1.128e-02  4.210e+02   0.422  0.67318
vowelOW     -2.224e-02  8.509e-03  2.450e+02  -2.614  0.00950 **
vowelUH     -4.951e-02  1.017e-02  2.490e+02  -4.866 2.02e-06 ***
vowelUW     -2.083e-02  1.021e-02  3.070e+02  -2.040  0.04218 *
```

Overall, participants' vowels differed by about 98 milliseconds in duration from the average vowel durations of the confederate ($\beta = 87.3$ msec, SE = 14.8 msec, $p < 0.001$). There was a significant positive effect of F1, indicating that participants differed more in low vowels than in high vowels ($\beta = 1.8$ msec, SE = 0.8 msec, $p < 0.05$), and a negative effect of F2, indicating that participants differed more in vowel duration from the confederate in back vowels than in front vowels ($\beta = 2.4$

msec, SE = 0.9 msec, $p < 0.01$). Participants diverged in average vowel length more in the middle of the experiment than at the end, by about 5 milliseconds ($\beta = 4.99$ msec, SE = 2.31 msec, $p < 0.05$); this is likely because the confederate had a harder task at the end of the experiment in deciding which forms to use during their turns, and so was speaking more slowly than the participants. There were also significant differences in the difference in duration of different vowels: participants' /i/ vowels were 20 to 50 milliseconds farther away from the confederate's mean /i/ duration than their /ɪ ɛ oʊ ʊ u/ vowels were from the confederate's respective vowels.

3.3.2.1.2 Convergence in vowel formants

To measure convergence in vowel formants, I normalized the first two formant frequencies for each vowel using the Bark scale and then measured the Euclidean distance in F1/F2 vowel space from the confederate's mean formant values for that vowel during the same timespan of the experiment. Taking subjects' vowels from the beginning and the end of the experiment, I calculated the difference in distance from the confederate's mean formant values for those two phases of the experiment. For this analysis I used the participants' first and last five primary stressed tokens of the vowels /æ ɪ oʊ/, and their first and last five unstressed tokens of /ə/. These vowels were chosen because they are far apart acoustically, and because there were a substantial number of tokens for each in the speech of most of the participants. For the confederate's vowels, I used all tokens of the vowel that were produced over the same timespan within the experiment. For example, if a participant produced their fifth stressed /ɪ/ six minutes into the game portion of the experiment, I compared these productions of /ɪ/ to average formant values of all of the tokens of stressed /ɪ/ that the confederated had produced within the first six minutes of the game. Due to the sparseness of the data, vowels' consonantal context could not be taken into account.

In Table 3.10, a negative sign indicates convergence (as mean Euclidean distance shortened over the experiment). Overall, the mean difference-in-distance across all subjects and vowels was –0.003 Bark (SD = 1.087). Across subjects, none of the four vowels under investigation had a distribution that was significantly different from zero. None of the subjects had a mean difference-in-distance across all of the vowels that was significantly convergent.

Table 3.10: Difference-in-distance of F1/F2 in vowels

Subject	/æ/	/ə/	/ɪ/	/oʊ/
S08	0.823	0.083	0.380	-0.008
S10	0.204	-0.105	-0.710	0.403
S11	-0.154	-0.342	0.657	0.280
S16	0.169	-0.151	0.320	1.141
S18	0.298	1.348	NA	0.556
S19	-0.691	0.416	0.379	0.895
S20	0.475	0.901	NA	0.122
S21	-0.041	-0.423	0.160	1.007
S22	-0.383	0.353	NA	NA
S23	-0.021	-0.185	0.160	0.621
S24	-0.285	-0.651	NA	NA
S25	0.020	0.480	NA	NA
S26	-0.130	0.352	-0.647	-1.086
S27	0.400	0.138	-0.227	-0.115
S31	-0.124	0.337	-0.269	-0.321
S32	0.439	0.289	NA	0.417
S33	-0.765	0.600	NA	NA
S34	-0.106	-0.400	-0.496	0.436
S35	-0.127	-0.164	-1.047	-0.188
S36	0.081	-0.625	0.150	-0.507
S37	-0.766	-0.039	NA	-0.040
S38	-0.257	-0.251	NA	-1.137
S39	0.326	0.570	-0.113	-1.095
S40	-0.299	-0.133	0.288	NA
S42	-0.336	-0.579	-0.202	-0.426
S43	0.140	-0.449	0.570	1.119
S44	-0.050	-1.075	-0.288	-0.621

In order to look at the effects of power on the distance between participants' vowels and the confederate's vowels in F1/F2 Bark space, I created a mixed-effects linear regression model using the lme4 package (Bates et al. 2016) in R (R Core Team 2016). The outcome variable was the Euclidean distance between each of the participants' vowels and the confederate's mean Bark-transformed F1 and F2 for that vowel within the same timespan of the experiment. Independent variables were time (either the beginning, middle, or end of the experiment), sex, vowel identity (with IY as the reference level), vowel duration, and the interaction between time and vowel identity as main effects; and subject and word as random effects. The interaction between time and vowel, if significant, would capture the possibility that convergence distance may be vowel-specific, as previous studies have indicated (Babel 2012). Vowel duration was not a significant or informative predictor, so I removed it from further analysis. The interaction between time and vowel, although significant at the $p < 0.05$ threshold, reduced the model's informativity; I removed the interaction effect from the model but retained the main effect. For each of the six power factors in consideration (the five empowerment factors as well as the judges' ratings), I created a separate model adding that factor, plus its interaction with time, as predictors. None of these

additions improved the model, and in only one case (SF3: *community activism and autonomy*) did any of these additional predictors have an effect, in interaction with time. Table 3.11 shows the model without any power factors as predictors, and Table 3.12 shows the model with SF3 as a predictor; *p*-values were supplied by the lmerTest package (Kuznetsova et al. 2016).

In the model with no power predictors (shown in Table 3.11), participants' vowels at the end of the experiment were farther from the confederate's vowels in the middle of the experiment than at the beginning (β = 0.07 Bark, SE = 0.03 Bark, $p < 0.05$). This may be due to greater variability in the middle phase of the experiment, which included everything that wasn't the generally comparatively short beginning and end phases. Male participants' vowels were closer to those of the confederate, who was male (β = –0.77 Bark, SE = 0.07 Bark, $p < 0.001$). Participants' vowels differed greatly in distance overall, with /i/ being the most different and /ə˞/ being the closest.

Table 3.11: Vowel F1/F2 distance model (no power predictors)

```
Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.05243  0.2290
 subject  (Intercept) 0.02388  0.1545
 Residual             0.47419  0.6886
Number of obs: 5957, groups:  word, 464; subject, 27

Fixed effects:
             Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   2.30048    0.08512  192.00000  27.026  < 2e-16 ***
TimeMid       0.07142    0.02786 5949.00000   2.564 0.010375 *
TimeEnd      -0.02999    0.02766 5930.00000  -1.084 0.278293
SexM         -0.76759    0.06643   26.00000 -11.555 1.10e-11 ***
vowelIH      -0.57817    0.09421  250.00000  -6.137 3.27e-09 ***
vowelEY      -0.30794    0.09406  201.00000  -3.274 0.001248 **
vowelEH      -0.58558    0.09347  218.00000  -6.265 1.97e-09 ***
vowelAE      -0.53796    0.09165  166.00000  -5.870 2.31e-08 ***
vowelAY      -0.33957    0.09954  187.00000  -3.411 0.000793 ***
vowelER      -0.81350    0.15082  532.00000  -5.394 1.04e-07 ***
vowelAW      -0.51731    0.13255  221.00000  -3.903 0.000126 ***
vowelAA      -0.52244    0.10032  240.00000  -5.208 4.11e-07 ***
vowelAH      -0.56212    0.10168  230.00000  -5.528 8.75e-08 ***
vowelAO      -0.18434    0.12527  350.00000  -1.472 0.142027
vowelOW      -0.31001    0.09469  199.00000  -3.274 0.001250 **
vowelUH      -0.76418    0.11518  215.00000  -6.635 2.61e-10 ***
vowelUW      -0.24689    0.11620  275.00000  -2.125 0.034496 *
```

Table 3.12: Vowel F1/F2 distance predicted by community activism and autonomy

```
Random effects:
 Groups   Name        Variance Std.Dev.
 word     (Intercept) 0.05188  0.2278
 subject  (Intercept) 0.02388  0.1545
 Residual             0.47387  0.6884
Number of obs: 5957, groups:  word, 464; subject, 27

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)    2.22517    0.08698  213.00000  25.583  < 2e-16 ***
SF3           -0.04525    0.04032   56.00000  -1.122 0.266555
TimeMid        0.07663    0.02793 5951.00000   2.743 0.006100 **
TimeEnd        0.04525    0.03564 5876.00000   1.270 0.204229
SexM          -0.76247    0.06976   26.00000 -10.929 3.47e-11 ***
vowelIH       -0.57921    0.09396  250.00000  -6.164 2.82e-09 ***
vowelEY       -0.30829    0.09378  200.00000  -3.288 0.001194 **
vowelEH       -0.58738    0.09320  217.00000  -6.302 1.62e-09 ***
vowelAE       -0.54087    0.09136  165.00000  -5.920 1.81e-08 ***
vowelAY       -0.34114    0.09924  186.00000  -3.438 0.000724 ***
vowelER       -0.81821    0.15055  533.00000  -5.435 8.36e-08 ***
vowelAW       -0.51362    0.13218  221.00000  -3.886 0.000135 ***
vowelAA       -0.52546    0.10006  239.00000  -5.251 3.33e-07 ***
vowelAH       -0.56193    0.10142  229.00000  -5.541 8.23e-08 ***
vowelAO       -0.19141    0.12501  350.00000  -1.531 0.126634
vowelOW       -0.31013    0.09441  198.00000  -3.285 0.001205 **
vowelUH       -0.76898    0.11486  213.00000  -6.695 1.88e-10 ***
vowelUW       -0.24374    0.11590  274.00000  -2.103 0.036380 *
SF3:TimeM      0.06289    0.02727 5941.00000   2.307 0.021116 *
SF3:TimeE      0.05203    0.03522 5913.00000   1.477 0.139648
```

For the model adding SF3, there was a positive interaction between SF3 and time, such that participants with a higher autonomy rating had vowels that were farther away from those of the confederate at later points in the experiment by 0.05 to 0.06 Bark (beginning to middle $\beta = 0.063$ Bark, SE = 0.027 Bark, $p < 0.05$; beginning to end $\beta = 0.052$ Bark, SE = 0.035 Bark, *n.s.*). All other effects were at approximately the same magnitude as in the equivalent model without SF3 as a predictor.

3.3.2.2 Analysis: Lexical convergence

Because of the difficulty inherent in determining convergent behavior without previous divergent behavior having been established, I did not analyze the first label used by a participant for a given referent, even if the confederate had already used a label for that referent.
If Form B differed from Form A, then Form C was coded as a convergence locus[22]. Example: the participant asks for a *plane* at Time A, and the confederate asks for an *airplane* at Time B. Form C is coded as a convergence locus. The locus is coded as a match if it is different from Form A (e.g. *airplane, aeroplane*), and as a mismatch otherwise (e.g. *plane*).

Additionally, sometimes the confederate used multiple different labels for an object in the same turn, presumably due to the difficult nature of his task. In these cases, Form C was coded as a match only if it was different from Form A, and as a mismatch only if it did not match either label used by the confederate at Time B.

Two of the objects had target labels that could be considered variant pronunciations rather than different lexemes. VASE's two target labels were as /vɛɪs/ and /vɑz/; ENVELOPE was to be referred to as either /ɛ/nvelope or /ɑ/nvelope. Only two participants showed any variation between the two pronunciations of VASE. Even more starkly, only one participant showed any variation between the two pronunciations of ENVELOPE, and that was in a false start. Because of this extremely low rate and the different nature of these two items, they were excluded from the analysis of lexical convergence.

In order to determine whether the absolute frequency of a given form determined the likelihood of convergence to that form, I looked at the correlation between the difference between the log word frequency of Form A and that of Form B (FreqB – FreqA = FreqDiff), and the log word frequency of Form C (FreqC). There was a slightly negative correlation that was not significant (–0.03, $p > 0.1$). I then created a logistic regression model with probability of lexical convergence as the outcome variable, a random effect of Subject, main effects of FreqA and FreqDiff, and an interaction effect between the two. Frequencies were compiled from the SUBTLEX corpus (Brysbaert & New 2009). For the purposes of the present analysis, accommodation and convergence were conflated such that any change in the choice of variable used by the subject, whether the result exactly matched the variable used by the confederate or not, was considered convergence. Multiple-word responses (e.g., *baby carriage*, *bee-wasp*) were discarded. Each convergence locus was then counted as an observation ($n = 194$), with the outcome variable being a two-level factor with a value of either N (for no change/no convergence) or Y (for convergence). The best-fit model did not include FreqA as a predictor, and FreqDiff did not prove to be a significant predictor of convergence, although there was a general tendency for greater convergence toward more common labels. The coefficient table of the resulting model is shown in Table 3.13. Overall, there was a 47.4% incidence of lexical convergence (at 92 out of 194 convergence loci).

22. There were three instances in the corpus where the participant changed their label to match a label the confederate had used prior to the participant's first label. Example: The confederate asks for a *wasp*, and then the participant asks for a *bee*, and then the confederate asks for a *bee*, and then the participant asks for a *wasp*. These three instances were not considered to be convergence loci.

Table 3.13: Lexical convergence predicted by frequency of forms used

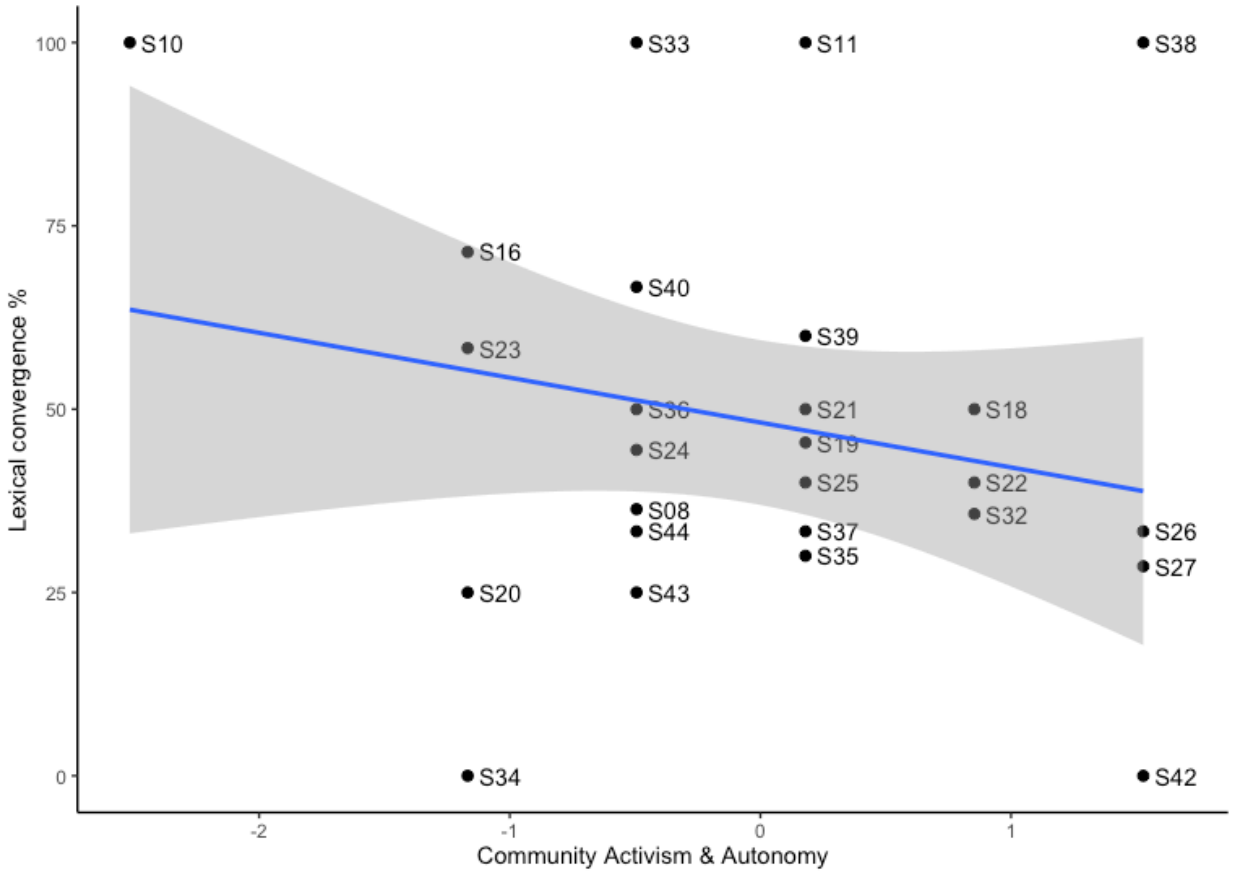
```
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.2916 -0.8787 -0.7331  1.0608  1.4011

Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.03783  0.1945
Number of obs: 148, groups:  Subject, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.08633    0.19143  -0.451    0.652
FreqDiff     0.34577    0.21100   1.639    0.101

Correlation of Fixed Effects:
         (Intr)
FreqDiff 0.416
```

For each of the six power factors in consideration (the five empowerment factors as well as the judges' rating), I created a separate model adding that factor as a predictor. In only one case (SF3: *community activism and autonomy*) did any of these additional predictors have an effect. Table 3.14 shows the model with SF3 added as a predictor. There was a slight negative effect of SF3, such that participants with higher reported autonomy ratings were less likely to evince lexical convergence ($p < 0.05$). Figure 3.4 shows participants' proportion of lexical convergence as a function of their SF3 rating, along with the best-fit line described in Table 3.14.

Table 3.14: Lexical convergence predicted by frequency and community activism/autonomy

```
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.5917 -0.8566 -0.6476  1.0217  1.6188

Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0        0
Number of obs: 148, groups:  Subject, 25

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.08214    0.18750  -0.438   0.6613
SF3         -0.41514    0.20598  -2.015   0.0439 *
FreqDiff     0.33707    0.21221   1.588   0.1122

Correlation of Fixed Effects:
         (Intr) SF3
SF3      -0.003
FreqDiff  0.429  0.005
```

Figure 3.4: Community activism & autonomy and lexical convergence



3.3.2.3 Analysis: Syntactic convergence

In order to look at syntactic convergence, I had several reasons for using the alternation between requests and statements. First, these constructions fit easily into the type of task I was looking for, in which objects are to be named repeatedly and in turn. Second, the alternation is clearly syntactic but not exclusively so. Within the context of this experiment the pragmatic aspects of the alternation between these constructions are hopefully minimized, making this experimental paradigm an ideal one in which to investigate the alternation. (It might not provide the clearest results when comparing syntactic to lexical and phonetic convergence, however.)

However, there is also an obvious potential confound in the data. The differences between requests and statements are partially related to interlocutor roles and politeness (Holtgraves 1986). As such, we may expect participants who see themselves as being in high-power positions in the interactions to be less consciously polite, and so to use statements. Conversely, we may expect low-power participants to rigorously use requests, as they are generally more polite.

Unlike phonetic convergence and like lexical convergence, syntactic convergence deals with discrete differences. Previous studies of syntactic convergence (or syntactic co-ordination as described by Garrod & Anderson 1987; Branigan et al. 2000) have scored every instance of a

construction for co-ordination, regardless of whether the construction is a change from the previous construction by that same speaker. As people are more likely to repeat structures they themselves have used (Garrod & Anderson 1987, Branigan et al. 2000), this kind of counting is confounded. So I will only look at places where the confederate uses a different structure from the participant, and examine the participant's following production.

In a control trial with two naive participants, every single turn of the game except for one used the syntactic template *"Can I have your _____?"*. In other words, after the first player set the template for the interaction, both participants adhered to it almost without exception. The single exception to this template, *"Then I'll need your rabbit",* came when it was clear from context that the utterer was intending a specific pragmatic message which was overlaid onto the otherwise-rote task. Because of this strong regularity, I interpret *any* deviation on the part of a participant as constituting syntactic accommodation. However, not all deviation should be considered syntactic *convergence*.

The confederate was therefore behaving unusually; people don't generally vary their utterances, let alone with the frequency at which the confederate did. We might therefore expect participants to have thought the confederate was strange and distanced themselves from him. While none of the participants reported noticing anything strange about the confederate's speech pattern, several seemed to register it after the purpose of the study was explained.

In order to determine whether there was a general difference in the likelihood of convergence toward requests compared to convergence toward other utterance forms, I created a logistic regression model predicting convergence as a function of the confederate's utterance form (FormB), with a random effect of Subject. The coefficient table of the resulting model is shown in Table 3.15. There was indeed a significant effect of utterance form, such that participants were more likely to converge toward requests than statements, imperatives, or other utterance forms ($p < 0.001$). For the purposes of the present analysis, accommodation and convergence were conflated such that any change in the choice of utterance form used by the participant, whether the result exactly matched the utterance form used by the confederate or not, was considered convergence. Each convergence locus was then counted as an observation ($n = 462$), with the outcome variable being a two-level factor with a value of either A (for no change/no convergence) or B (for convergence toward Form B). Overall, there was a 26.0% incidence of accommodation to utterance type (at 120 out of 462 convergence loci).

Table 3.15: Syntactic convergence predicted by type of utterance used

```
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 2.753    1.659
Number of obs: 462, groups:  Subject, 27

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.9665     0.4626  -4.251 2.13e-05 ***
FormBoth      2.2272     1.1681   1.907   0.0566 .
FormBreq      3.7008     0.6390   5.792 6.97e-09 ***
FormBsta     -0.3420     0.3692  -0.926   0.3543

Correlation of Fixed Effects:
         (Intr) FrmBth FrmBrq
FormBoth -0.147
FormBreq -0.375  0.107
FormBsta -0.509  0.160  0.359
```

For each of the six power factors in consideration (the five empowerment factors as well as the judges' rating), I created a separate model adding that factor as a predictor. Again, only community activism and autonomy (SF3) was close to significant; Table 3.16 shows the model with SF3 added as a predictor. There was a slight tendency of participants with higher reported SF3 ratings to evince less convergence toward utterance type, although this tendency was not significant at the $p < 0.05$ threshold ($p > 0.05$). Figure 3.5 shows participants' proportion of convergence to utterance type as a function of their SF3 rating, along with the best-fit line described in Table 3.16.

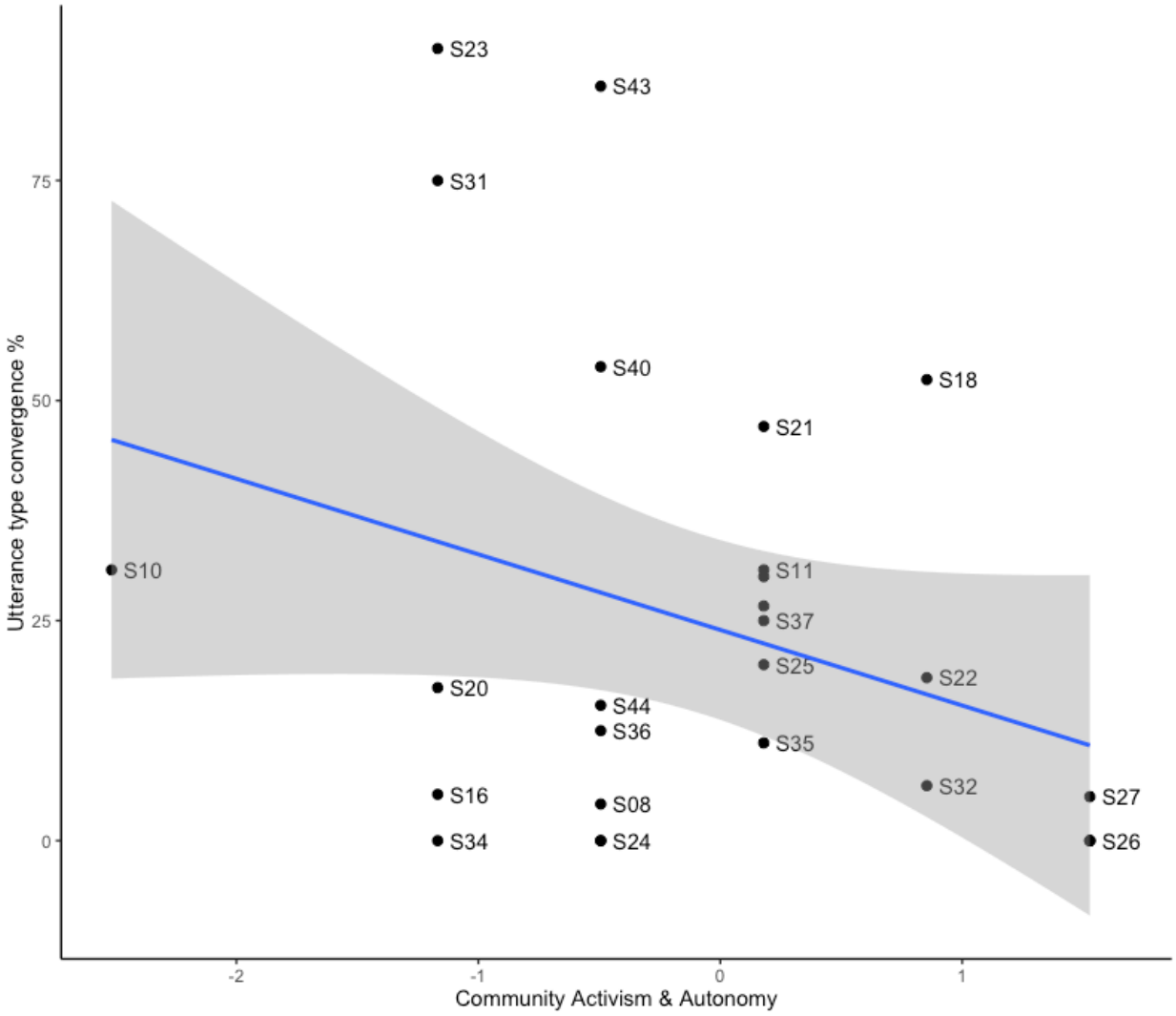
Table 3.16: Convergence to utterance predicted by type of utterance and community activism/autonomy

```
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 2.48     1.575
Number of obs: 462, groups:  Subject, 27

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.0420     0.4553  -4.485 7.30e-06 ***
SF3          -0.6294     0.3813  -1.651   0.0988 .
FormBoth      2.2711     1.1630   1.953   0.0508 .
FormBreq      3.6997     0.6365   5.812 6.16e-09 ***
FormBsta     -0.3031     0.3694  -0.821   0.4118

Correlation of Fixed Effects:
         (Intr) SF3    FrmBoth FrmBrq
SF3       0.177
FormBoth -0.154 -0.035
FormBreq -0.382 -0.053  0.109
FormBsta -0.515 -0.036  0.164   0.359
```

Table 3.17: Correlation table of participant convergence rates across levels

	Lexeme Choice	Utterance Choice	Vowel Durations
Utterance Choice	0.104		
Vowel Durations	-0.164	-0.122	
Vowel Formants	-0.147	-0.189	0.377

Figure 3.5: Community activism & autonomy and convergence to utterance type



3.3.2.4 Analysis: Across levels

To address the question of whether convergence occurs at the same rate at different levels of linguistic structure, I took the results of the last four analyses and looked at correlations between them, by subject. Table 3.17 shows the correlations between average vowel formant difference-in-distance measurements (in Bark), absolute vowel duration differences (in milliseconds), lexeme convergence (as a proportion), and utterance type convergence (as a proportion), by subject. There was a weak positive correlation between the two phonetic-level variables (r = 0.377, n.s.), and a weak positive correlation between the two higher-level variables (r = 0.104, n.s.), but weak negative correlations between the lower- and higher-level variables. None of these correlations were significant. I also created a logistic regression model with probability of convergence as the outcome variable and a main effect of level of linguistic structure. For this analysis, I only used

categorical data (i.e., no phonetic data). As in the previous analyses in this chapter, convergence was defined as a change in the choice of variable used by the subject, whether the result exactly matched the variable used by the confederate or not. The resulting model is shown in Table 3.18.

Table 3.18: Convergence as a function of level of linguistic structure

```
Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.736    0.8579
Number of obs: 1648, groups:  Subject, 27

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.2229     0.2032  -6.018 1.76e-09 ***
Levelaux      0.4798     0.1778   2.699  0.00696 **
Leveldet      0.3853     0.2259   1.706  0.08800 .
Levellex      1.0862     0.1955   5.556 2.76e-08 ***
Levelphonol   0.3101     0.5975   0.519  0.60374
Levelverb     0.3884     0.1509   2.574  0.01007 *

Correlation of Fixed Effects:
            (Intr) Levelaux Leveldet Levellex Lvelphon
Levelaux    -0.376
Leveldet    -0.288    0.338
Levellex    -0.333    0.392    0.293
Levelphonol -0.104    0.127    0.092    0.121
Levelverb   -0.431    0.494    0.391    0.447    0.146
```

For this model, *Level* = level of linguistic structure. The levels for this predictor are *aux* = syntactic: presence and choice of auxiliary; *det* = syntactic: choice of determiner; *lex* = lexical: choice of lexeme; *phonol* = phonological: choice of pronunciation; *type* = syntactic: type of utterance (statement vs. request); *verb* = syntactic: choice of verb.

There was a moderate positive correlation across all levels of linguistic structure under investigation (excepting the phonological level); however, there were significant differences in the global proportion of convergence at different levels ($p < 0.001$). Only lexical convergence has a greater than 50% likelihood of occurring, across subjects (87.2%, $p < 0.001$) according to this model. For each of the six power factors in consideration (the five empowerment factors as well as the judges' rating), I created a separate model adding that factor as a predictor. None of these potential factors were significant predictors at the $p < 0.05$ threshold or improved the model.

3.4 Discussion and conclusions

Unlike most studies examining phonetic convergence in natural speech, the current study is looking at convergence of an individual toward their interlocutor, rather than global

convergence between dyads; unlike most studies examining phonetic convergence on an individual basis, the current study is looking at natural speech. As such, neither a by-token approach nor an overall convergence measure fits this data.

Hypotheses. The hypotheses under consideration are listed again in Table 3.19. The first set of hypotheses concerned the congruity of convergence behavior across levels of linguistic structure.

H1–H3: There was no strong pattern found regarding correlated convergence across levels of structure, as seen in Table 3.17. If anything, this lack of a pattern is evidence that there is no community-wide relationship between convergence rates for different structures, militating against H1 and H3 but not against H2. This finding argues against the interconnectedness of different levels of representation that is proposed in Interactive Alignment Theory (Pickering and Garrod 2004).

Table 3.19: Hypotheses under consideration for game experiment

H1: People will converge to phonetic, lexical, and syntactic levels of structure at a similar rate.

H2: People will converge to each level of structure contingent on its role in the conversational interaction.

H3: Phonetic and lexical convergence will occur at a similar rate, which may differ from syntactic convergence.

H4: People with higher power will converge less.

H5: People with lower power will converge less.

H4–H5: The second set of hypotheses concerned the effects of personal power on convergence. The findings of this study were inconclusive with regard to this question. However, it is intriguing that the empowerment factor representing community activism and autonomy repeatedly approached significance as a predictor, for convergence to each of vowel formants, lexeme choice, and utterance choice. It is plausible that there is an inverse relationship between a person's sense of autonomy and their attention to the low-level details of others' behavior. However, this interpretation is independent from the aspects of power that form the impetus for the current analysis, although it seems to warrant further investigation.

Overparticular categorization. In the context of the account I am sketching in this dissertation, non-activation of potentially relevant categories might be ascribed to working memory or inhibitory considerations. However, this may not be an active process of suppression. Rather than the activation of particular units of representation being reduced, under the categorization schema account those units are never activated in the first place. In this account, inhibition of imitation comes about when a speaker's categorization schema is overparticularized: the new situational context-level category is defined too narrowly or rigidly for incoming speech to fit in it.

Levels of representation: I want to note here that the term *level of representation* is a theoretical one. Pickering and Garrod (2013) take the existence and independence of different levels of linguistic representation as a given: "Unlike many other forms of action and perception,

language processing is clearly structured, incorporating well-defined levels of linguistic representation such as semantics, syntax, and phonology. Thus, our accounts also include such structure" (Pickering and Garrod 2013:332). While it is reasonable to suppose that different processes underlie linguistic chunks of different sizes, there are no clear divisions of structure at any point along the continuum from sounds to phones to phonemes to syllables to morphemes to words to phrases to utterances. It seems reasonable to assert that the mind uses different strategies for handling differently-sized chunks of language; however, it is not clear that these strategies are common across individuals, or even necessarily across contexts.

Chapter 4: Consequences of convergence

4.0 Introduction

The initiation of sound change requires the emergence of innovated variants. One possible source of innovated variants is *non-faithful transference* of linguistic features during accommodation to a received speech signal. By non-faithful transference I mean a lack of phonetic fidelity between the signal perceived and the signal produced; I do not mean to invoke the idea of phonological faithfulness. If the subphonemic cues associated with a phonological contrast are realized differently in an accommodating speaker's speech than in both the speaker's previous speech and in the signal to which they are converging, the resulting difference in phonetic realization may constitute a new variation.

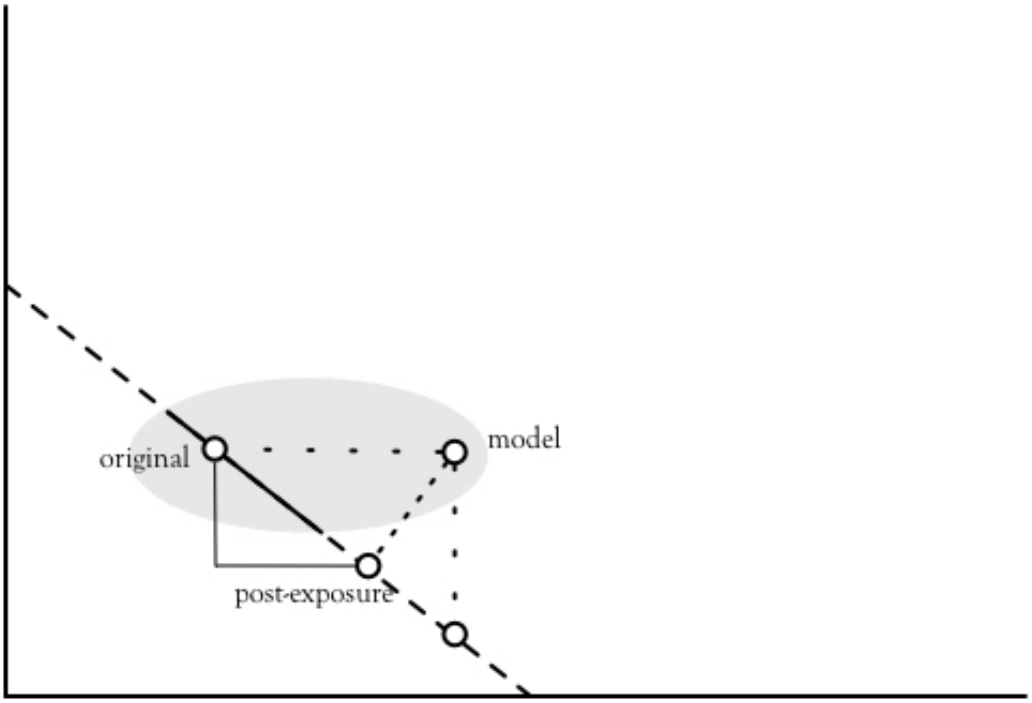
In this chapter I argue that accommodation must logically lead to this sort of difference, at least for some speakers along some phonetic dimensions. I then report the results of a study demonstrating that such phonetic divergence does happen in a laboratory environment, indicating that accommodation is, paradoxically, a potential source of new phonetic variants. I conclude by situating these findings within extant work on sound change. To the extent that populations of individuals evince varying degrees of non-faithful transference of these cues, this study provides evidence for a potential approach to the actuation problem (Weinreich et al. 1968).

4.1 Non-faithful accommodation

Delvaux & Soquet (2007) hypothesize that imitation in speech explains both the stability of phonetic realizations within a speech community, and the potential for change to a community's speech norms via the transmission of new realizations. However, it may be possible to impute not only the transmission of sound change, but also the *actuation* of sound change, to the effects of accommodation. In other words, can accommodation alone — in the abstract, without appeal to supplemental considerations — directly result in the introduction of innovative forms? The wide-ranging nature and robustness of accommodation effects suggests that this idea is worth considering.

If the incorporation of phonetic episodes leads to a new phonetic target that cannot be faithfully produced, any compromise creates the possibility of divergence in the features that are compromised. I am calling this possibility *antagonistic* accommodation, in the sense that the cues that are variously convergent and divergent are acting against each other. Antagonistic accommodation is a likely domain within which sound change actuation may happen, due to the recoupling of gestures in new ways due to the pressures placed on an individual's speech system by either physiological or psychological/phonological factors. Figure 4.1 shows a schematization of this potentiality.

Figure 4.1: Antagonistic accommodation leading to innovated variation



The *x* and *y*-axes represent two cues to a phonetic feature that are in a physiologically or psychologically-induced antagonistic relationship: within an individual's speech, *x* increases as *y* decreases, and vice versa. The shaded area is the extant area of variation within the speech community for the feature in question. The dotted line represents the (physiologically or phonologically) restricted range of variation the speaker has available for the two cues in question; the solid line represents the speaker's available range within extant variation. The point not on the dashed line represents received speech from a model/interlocutor; the dotted lines represent the shortest distances between the received speech and the speaker's available range of variation, depending on whether one, the other, or both cues are adjusted. The thin horizontal and vertical lines represent the amount of convergence in the *x* feature and the amount of divergence in the *y* feature, respectively.

For a concrete example, assume that the *x*-axis represents voice onset time (VOT) and the *y*-axis represents stop closure duration. The community as a whole has a relatively wide range of variation in VOT across speakers, but the range of closure duration values is more restricted. When the speaker adjusts their VOT in convergence to the model, the resulting token is outside of the community's extant variation; when the speaker adjusts only their stop closure duration, the result is within the established range of variation. However, in the event that a speaker adjusts both cues in order to approximate the model's variant as closely as possible along both axes, the resulting production is also outside of the community's extant variation. In converging to a greater degree in VOT, the speaker is diverging from the model in stop closure duration.

Unless accommodation can result in non-faithful transference of linguistic features from the received signal, it cannot explain the appearance of innovative forms within a population. However, if it *can* — if accommodation can result in the appearance of entirely new linguistic

material — then accommodation is a possible source of permanent changes to the speech patterns of whole communities.

Previous studies have shown that speakers do evince non-faithful accommodation. Nielsen (2011) looked at whether accommodation can lead to a phonological generalization of received subphonemic features. Participants were exposed to model speech with extended VOTs exclusively in words beginning with /p/, and then recorded saying words beginning with both /p/ and /k/. Nielsen found that VOT was indeed generalized to /k/ in this fashion, although new words with initial /p/ evinced a higher degree of VOT adjustment than words with initial /k/. Additionally, words with a lower lexical frequency showed a higher degree of VOT adjustment than more frequent words. Nielsen's study established that non-faithful transference is possible in accommodation. However, the particular type of transference investigated by Nielsen does not result in a new pattern of phonetic material, as /k/ with extended VOT is found in English in general, and there was no VOT increase evinced in segments without such attested extensions, for instance /m/.

As of yet it has not been demonstrated whether accommodation can result in changes that were not present at all in the speech signal being accommodated toward. If it cannot — if all accommodation results in faithful changes to the received signal — the process can only result in the mingling of already extant linguistic features in the speech profiles of disparate individuals within a speech community or population. Similarly, changes that differ only in degree will not affect the constitution of a category unless those changes result in the extension of the category's boundaries. If a category has two non-overlapping areas in phonetic space associated with it, it is possible for a change in degree to lie between those areas. But for a contiguous and convex category, adjustment toward an already-extant variant will necessarily lie within the category as previously defined. Adjustment *away from* an already-extant variant, on the other hand, may well extend category boundaries.

Antagonistic accommodation is not the only type of accommodative behavior that is a potential source of innovative material. While convergence toward the interlocutor will not result in the introduction of new variations into the language, *hyperconvergence* — accommodation toward *and past* the speech of the interlocutor (Giles 1971) — may result in new variation. To the extent that they are attributable to automatic physiologically determined imitation (Gentilucci & Bernardis 2007), both antagonistic accommodation and hyperconvergence provide speaker-level sources of community-level innovation, referred to by Chang (2012) as "phonetic drift" (repurposed from Sapir 1921). It is also conceivable that a change stemming from divergence could conceivably result in new variation, to wit the extension of a category away from the examples provided by an interlocutor. As mentioned, however, it has not been demonstrated that such changes persist beyond the interaction in which they were instantiated.

4.2 The current study

The current study addresses the question of whether antagonistic accommodation happens by looking at *coincident cues* (also called redundant cues): features that tend to coincide with a given linguistic feature, signaling or reinforcing the identity of that feature. This type of reinforcement of information-laden units is exceedingly common on all levels of linguistic structure. In this paper, I will look at phonetic reinforcement, namely coincident cues to lengthened VOT in voiceless stops

in English. The study discussed here examines the degree of automaticity in accommodation to three cues associated with VOT — duration of closure of the stop in question, as well as both the initial F0 and total duration of the vowel following the stop — when only the VOT is experimentally manipulated.

The current experiment looks at coincident cues of VOT for several reasons. VOT is generally held to be the most information-laden feature marking the phonological distinction between voiced and voiceless stops in English, as phonologically voiced stops are typically realized as voiceless, especially in onsets of stressed syllables. Previous studies have shown that the length of VOT of English voiceless stops can be manipulated in phonetic accommodation (Fowler et al. 2003; Shockley et al. 2004). VOT is relatively easy to measure with consistency, due to the clearly definable features signifying its onset (stop release) and endpoint (glottal pulse). Finally, there are many coincident cues of VOT in English with varying degrees of automaticity, three of which are under investigation in this study.

Stop closure duration: VOT is inversely correlated with stop closure duration (Lehiste 1970, Boucher 2002). For a given speaker, overall timing relations are such that the overall stop duration will remain approximately constant: as such, an increase in VOT is concurrent with a decrease in closure duration, and vice versa.

Initial F0 of following vowel: Higher onset vowel F0 is a perceptual cue for voiceless stops in English (House & Fairbanks 1953). However, there is no direct correlation between VOT and F0 onset in production (Hombert 1976), and there are crosslinguistic differences in the direction of correlation between F0 and VOT across stop categories within languages (Kingston & Diehl 1994). Speakers of French are known to manipulate vowel F0 in accommodation (Delvaux & Soquet 2007). Among a population of English speakers, an increase in VOT may be expected to be accompanied by an increase in onset vowel F0; however, as this relationship may not be an automatic one, this expectation is a weak one.

Duration of following vowel: A longer VOT is correlated with a longer following vowel duration, due to general pressures for a fixed ratio of segment durations across speech rates (Boucher 2002). Vowel duration is known to be accommodated in other contexts than VOT lengthening (Giles et al. 1991).

I will use stop closure duration to illustrate how coincident cues might illustrate antagonistic accommodation. As previously discussed, stop closure duration is inversely correlated with VOT duration (Lehiste 1970, Boucher 2002). Given a model speech signal with lengthened VOT and average stop closure duration, different predictions are possible regarding the behavior of accommodating speakers as regards stop closure duration. (Assume for now that speakers have the same average closure duration as the model.) We might predict that speakers will shorten their stop closure duration in order to signal the VOT increase being targeted in accommodation to the VOT-adjusted speech signal. Alternatively, we may predict that speakers will increase their VOT but leave stop closure constant. It is not important at this point which of the two features is more likely to be adjusted; the predictions made in either case are formally equivalent in a discussion of the types of accommodation that are possible.

But what if speakers have a shorter average stop closure duration than the model signal, as well as a shorter VOT? In this case, convergence in both features would run counter to the tendency for their inverse correlation. In this instance we have a different set of predictions, shown in Table 4.1.

Table 4.1: Possible speaker adjustments for coincident cues across conditions (model with longer VOT and closure than speaker)

	VOT	Closure	Explanation	Result
1.	same	longer	Accommodation to closure, not to VOT	Convergence
2.	longer	same	Accommodation to VOT, not to closure	Convergence
3.	longer	longer	Convergence in both VOT and closure	Convergence
4.	longer	**shorter**	Convergence in VOT, coincident adjustment to closure	Antagonistic
5.	**shorter**	longer	Convergence in closure, coincident adjustment to VOT	Antagonistic

In strategies 1-3, all changes to VOT and stop closure are in the direction of phonetic convergence. However, the resulting adjustments are contrary to the general tendency observed for individual speakers as regards the relationship between the two features. In strategies 4–5, the speaker is exhibiting divergence of one of the two features in question, and convergence of the other. If one of the latter two predictions holds, the result is antagonistic accommodation, which can be attributed to restrictions in the speaker's timing relations.

In the event that speakers have a longer average stop closure duration than the model signal, as well as a shorter VOT, strategies 3–5 are functionally equivalent as far as predicted behavior. If speakers lengthen their VOT in convergence, coincident adjustment to closure duration will be in the same direction as direct accommodation would predict; if they shorten their closures, coincident adjustment to VOT would converge in length.

As the intent of the reported study was to investigate the existence of antagonistic accommodation, a model speaker was used with long closure duration as well as lengthened VOT, in order to make strategies 4 and 5 available to speakers. Given that previous studies have consistently found convergence in VOT, and the trade-off between VOT and stop closure is well established, it was predicted for this study that speakers would generally pursue strategy 4: they would diverge from the model in closure duration and converge in VOT.

Predictions for speakers' adjustment of vowel duration and onset F0 are contingent on their individual values for these cues. Given that the relationship between F0 and VOT is not clearly an automatic one, speakers are less likely to diverge in F0 as a physiologically inevitable byproduct of convergence in VOT. My prediction then is that they will either not adjust onset F0, or converge in vowel F0, but do so independently of their convergence in VOT. For vowel duration, coincident cue adjustment is predicted. However, some speakers will likely have longer, and others shorter, baseline vowels than the model, meaning that divergence in vowel duration is expected for some but not all speakers. It is also possible that some subjects might converge toward speaking rate or voice pitch as independent targets, not just as coincident cues of stop voicing.

As it is possible that individuals will use different strategies in accommodating to their interlocutor during the experiment, it is possible that each of these patterns will be evinced by different speakers. For example, some speakers may adjust onset vowel F0 while others do not. Among those who do, some may converge slightly toward the model talker, while other speakers exhibit hyperconvergence. Should such differences in accommodation be found, they may be attributable to social characteristics of the speakers. Yu (2013) delineates a set of predictions in this

vein, proposing a link between personality traits and the degree of compensation for coarticulation evinced in listening tasks. In order to investigate the possibility that a particular speaker's profile of social characteristics might in some way predict the ways in which coincident cues are adjusted in accommodation, the current experiment included a Big Five personality questionnaire (Saucier 1994, following Yu 2013), and a personal empowerment questionnaire (Rogers et al. 1997).[23]

4.3 Experiment

4.3.1 Methodology

The experiment consisted of three blocks: (1) baseline recording, (2) repetition of target exposure, and (3) post-exposure recording. Each session typically lasted twenty minutes. Participants were tested individually in a sound booth equipped with a PC, a microphone (AKG C3000), and headphones (AKG K240 Studio). The experimental stimuli were presented using a Python script. Prior to the baseline recording, participants filled out a questionnaire containing biographical data (age, sex, places of residence and languages spoken), a Big Five personality questionnaire, and a personal empowerment questionnaire.

In the baseline recording block, the words in the production list were visually presented on a monitor one at a time. Words persisted on the screen for 3 to 3.1 seconds with the interval in a uniformly random varying distribution, and were then replaced by the subsequent word with no break in between. Participants were given the following instruction: "You will see words. Please say them clearly and quickly." In the target exposure block, the participants were asked to repeat the words that they heard produced by the model talker over headphones. The instruction read: "Now you will hear words. Please repeat them clearly and quickly." No visual stimuli were presented in the exposure block. Finally, in the post-exposure recording block (which was identical to the baseline recording block), the participants were instructed to produce the words in the production list for a second time, providing a post-exposure recording. Across the three blocks, the words were presented in random order for each subject. Participants' tokens were digitally recorded into a computer at a sampling rate of 22,050 Hz.

The production list consisted of 120 words: 104 test words and 16 filler words (see Appendix B). For each word, the target segment was a voiceless stop in the onset of the word's stressed syllable. Among the test words, 49 had initial target voiceless stops (e.g., TENSION), 37 had initial schwa (e.g., ATTENTION), and 18 had initial unstressed syllables with onsets (e.g., DETENTION). Five of the words with initial targets and five of the words with initial schwa had glides following the stop (/k/ in all cases). The same word list was used in both the baseline and post-exposure conditions.

In order to forestall fatigue in participants, the repeat list was shorter than the production list, consisting of only 80 words: 65 test words and 15 filler words. Each word was repeated, resulting in a total of 160 list items. Among the test words, 38 had initial stops, 26 had initial schwa, and 1 had an initial unstressed syllable with an onset (MACAW). One of the words with an initial stop and one of the words with an initial schwa had glides following the stop (CHOIR and ACQUIT). One of the words in the listening list was not on the production list (PASTRY).

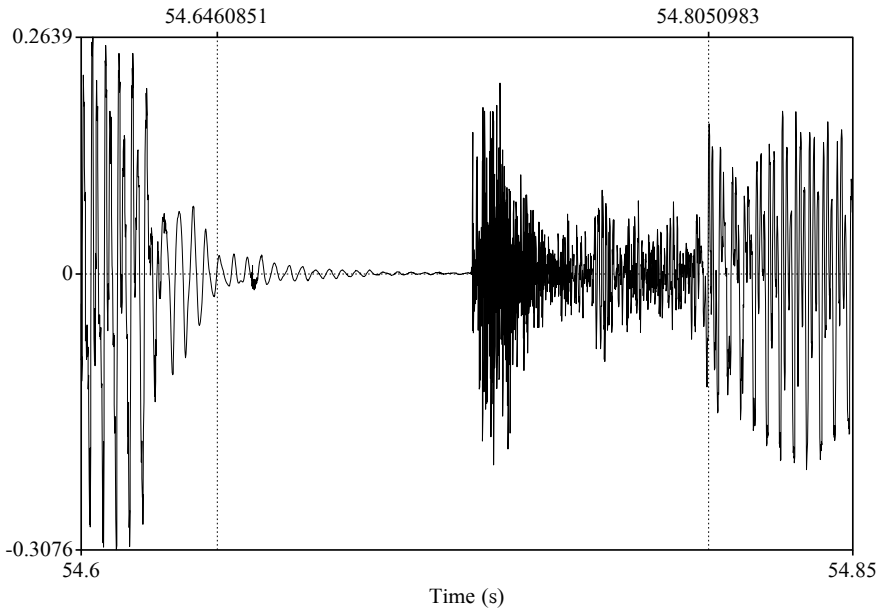
[23]It is also possible that differences in accommodation may be due to individual differences at the level of auditory perception; this study has nothing useful to contribute to such a line of inquiry.

A female American English speaker native to the San Francisco Bay Area (in which the experiment was conducted) served as the model talker, and was recorded saying each of the words in the listening list three times. The tokens were digitized at 44,100 Hz. The VOTs for the voiceless stops in the onsets of the stressed syllables for all three tokens of each word were artificially doubled using PSOLA resynthesis (Charpentier and Stella 1986) in Praat (Boersma & Weenink 2014), such that all parts of the consonant burst and aspiration were extended equally. Only the target segment was manipulated for each token, even if there were other onset voiceless stops in the word. The subjectively most natural token of each word post manipulation was used as a stimulus for the experiment. Overall mean VOT of stressed voiceless consonants in modeled speech was 154.08 ms, with standard deviation 28.52 ms. Mean closure duration for word-medial stops was 61.48 ms. Despite the model's unusually long VOT, in exit interviews participants described the model speaker as sounding natural, albeit hesitant.

Participants were compensated with US$5 and course credit. Of the thirty participants, thirteen did not meet the screening criteria for this study. Only three remaining participants were male; their data was excluded from analysis due to the lopsided population size and expected differences between sexes regarding onset vowel F0, one of the cues under investigation. The remaining fourteen were female, monolingual English speakers who gave permission to have their recordings used for analysis and discussion. Subsequently, VOTs and closure durations from the baseline and post-exposure blocks for these fourteen participants were measured from both waveforms and spectrograms using Praat (Boersma & Weenink 2014). Tokens that were pronounced with unexpected stress were excluded from subsequent analysis (n = 6 out of 2,912 target tokens). Due to the difficulties inherent in gauging the onset of stop closure in postpausal position, stop closure durations were only measured for word-medial stops. VOT was measured as the time between the onset of the release burst and the onset of periodic energy due to glottal pulse. Because periodicity was sometimes unclear, the trough before the first unequivocal period was taken as the onset.[24] Closure duration was measured as the time between the initial drop in intensity after periodicity had ceased in the preceding vowel and the onset of the release burst. Figure 4.2 shows an example of the measurement of overall stop duration (for the word ATTESTED).

[24]This determination has clear ramifications for the calculation of onset vowel F0.

Figure 4.2: Example of stop duration measurement for the word ATTESTED



4.3.2 Results across speakers

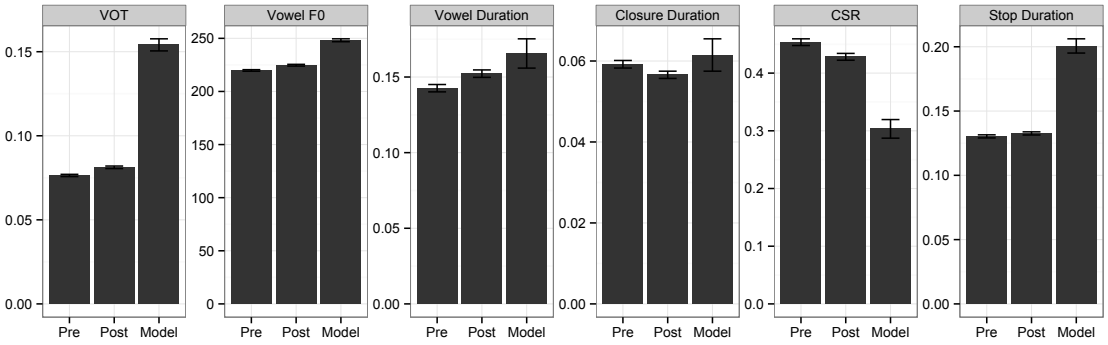
Table 4.2 shows a summary of the mean measurements across speakers in the pre-exposure and post-exposure conditions. Across speakers, average VOT increased from the pre-exposure condition to the post-exposure condition. Both vowel F0 and vowel duration also increased across speakers, converging toward the model talker's average for those features. Mean stop closure duration across speakers diverged from the model talker, decreasing despite the model talker's longer mean closure duration. However, the mean CSR (closure-stop ratio) across speakers converged toward the model talker's CSR. Additionally, the mean overall stop duration across speakers did not change significantly ($p = 0.158$). These data are illustrated in Figure 4.3.

Table 4.2: Global changes across conditions

	Condition		Model		ANOVA		Error *p*-values	
	Pre	Post	average	M	*F*	*p*-value	Speaker	Word
VOT (ms)	74.62	79.21	154.08	4.589	68.76	< 0.0001	0.107	0.54
Vowel F0 (Hz)	219.61	225.02	248.18	5.503	72.05	< 0.0001	0.517	0.997
Vowel duration (ms)	137.01	146.83	165.57	9.877	68.06	< 0.0001	0.454	0.838
VOT (ms)†	69.75	74.51	139.15	4.758	39.06	< 0.0001	0.0698	0.611
Closure duration (ms)†	56.74	**53.35**	61.48	3.32	18.48	< 0.0001	0.583	0.936
Closure-stop ratio (%)†	44.60	41.67	30.34	2.82	33.23	< 0.0001	0.123	0.58
Overall stop duration (ms)†	126.58	128.10	200.63	1.43	1.994	0.158	0.269	0.735

All ANOVAs included Condition as a predictor, and Speaker and Word as error terms.
***p*-values refer to significance of change from pre-exposure to post-exposure condition.**
†Does not include word-initial stops.

Figure 4.3: Global changes across conditions



Vowel F0 is in Hz, Closure-Stop Ratio is as a percentage, all other cues are in seconds.

Taken together, the divergence in absolute closure duration and lack of change in overall stop duration strongly suggest that VOT and closure duration are in an antagonistic relationship as coincident cues. Global convergence in CSR further indicates that divergence in closure duration is a side effect of convergence along other dimensions.

4.3.3 Results by speaker

Tables 4.3 and 4.4 show a summary of the mean measurements for each condition broken out by speaker.

Table 4.3: Individual changes across conditions (vowel cues)

	VOT (ms)		Vowel F0 (Hz)		Vowel duration (ms)	
Speaker	Pre	Post	Pre	Post	Pre	Post
Model	154.08 ms (28.52)		248.18 Hz (10.78)		165.57 ms (76.81)	
S15	76.39 (14.75)	91.50 (17.66)***	233.01 (17.51)	236.49 (22.94)	170.38 (77.23)	**184.07** (92.79)
S17	67.02 (15.20)	69.89 (18.09)	189.69 (20.02)	**189.28** (12.53)	107.32 (45.71)	113.12 (48.76)*
S20	63.11 (17.36)	**62.55** (14.89)	230.14 (9.82)	236.84 (11.88)***	152.04 (63.04)	155.72 (61.90)
S21	83.56 (19.38)	85.82 (21.50)	221.65 (29.05)	253.53 (17.25)***	144.73 (75.19)	151.31 (64.61)
S23	73.81 (16.37)	77.02 (18.50)	219.40 (18.77)	229.44 (14.77)***	158.02 (69.25)	160.85 (68.82)
S24	71.24 (14.51)	73.13 (19.34)	221.66 (19.84)	225.00 (23.06)	139.47 (63.71)	**138.05** (62.02)
S25	79.05 (13.23)	84.44 (16.07)***	201.70 (9.82)	202.40 (15.16)	103.37 (45.11)	111.14 (49.58)**
S28	70.57 (19.26)	**69.43** (18.21)	216.05 (17.19)	228.94 (10.24)***	108.16 (56.39)	137.32 (71.27)***
S30	80.95 (14.64)	82.56 (18.40)	229.71 (11.45)	235.12 (11.14)***	156.11 (61.64)	156.50 (59.56)
S32	74.33 (18.48)	79.70 (20.08)*	227.27 (12.79)	231.80 (14.40)**	118.92 (54.25)	131.63 (50.21)***
S35	92.35 (20.36)	100.41 (21.45)***	233.64 (10.03)	237.86 (13.12)	145.91 (68.97)	161.54 (77.80)***
S39	77.13 (20.27)	89.23 (18.87)***	205.47 (22.52)	210.11 (14.10)	134.44 (64.09)	155.10 (61.40)***
S40	80.47 (18.62)	83.63 (20.07)	233.94 (11.01)	235.00 (12.03)	158.95 (73.87)	178.46 (78.81)***
S42	54.42 (14.75)	60.23 (14.98)**	211.09 (19.22)	**198.29** 16.78)***	120.19 (50.62)	120.64 (51.78)

VOT is for all target stops. Features exhibiting divergence are bold and shaded. Features exhibiting hyperconvergence are in dashed boxes. Numbers in parentheses are standard deviations. Asterisks indicate significance levels of t-tests across conditions after Bonferroni correction. *$p < 0.05$, **$p < 0.01$, ***$p < 0.005$.

Table 4.4: Individual changes across conditions (word-medial stops)

	VOT (ms)		Closure (ms)	
Speaker	Pre	Post	Pre	Post
Model	139.15 ms (22.58)		61.48 ms (20.37)	
S15	72.47 (16.34)	85.45 (17.46)***	56.55 (13.70)	**47.78** (17.69)*
S17	60.62 (14.59)	64.10 (18.36)	52.21 (12.34)	**50.62** (13.59)
S20	62.43 (20.05)	62.91 (14.49)	49.72 (21.90)	55.20 (19.49)
S21	79.41 (17.17)	88.89 (21.16)*	55.05 (24.08)	**43.52** (16.70)**
S23	66.31 (14.14)	66.60 (14.20)	69.65 (22.93)	63.89 (16.17)
S24	67.83 (15.71)	68.15 (18.56)	54.37 (20.15)	**45.86** (15.22)*
S25	77.13 (12.80)	82.47 (15.47)*	61.81 (12.73)	**69.18** (13.33)***
S28	68.67 (21.29)	67.88 (19.99)	51.13 (15.07)	62.29 (11.69)***
S30	78.67 (14.01)	**76.73** (17.50)	44.49 (15.75)	**36.48** (8.27)**
S32	65.12 (15.07)	69.15 (14.92)	60.23 (16.68)	**53.63** (15.30)
S35	83.27 (16.05)	93.55 (20.83)***	66.67 (15.80)	49.73 (13.18)***
S39	69.78 (19.25)	83.69 (15.90)***	69.83 (18.80)	58.30 (17.15)***
S40	76.25 (19.53)	79.42 (18.92)	43.51 (16.05)	48.85 (15.76)
S42	50.89 (16.42)	56.97 (15.53)	58.76 (12.37)	61.37 (16.87)

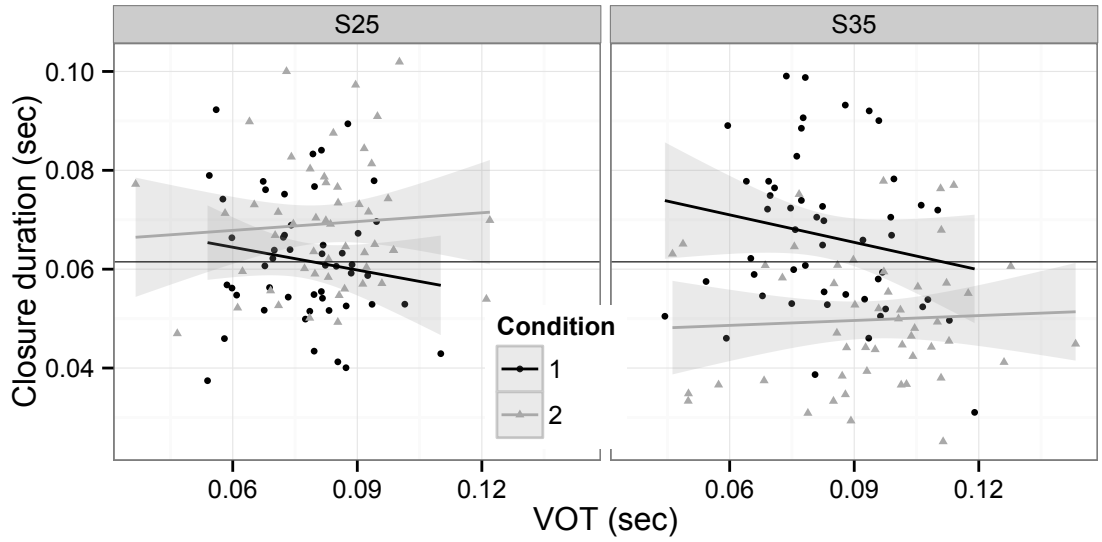
	CSR (%)		Stop duration (ms)	
Speaker	Pre	Post	Pre	Post
Model	30.34% (8.22)		200.63 ms (28.23)	
S15	43.88 (8.19)	35.29 (10.92)***	129.02 (20.55)	133.23 (24.27)
S17	46.44 (9.19)	44.40 (10.20)	112.83 (17.64)	114.73 (21.65)
S20	43.87 (15.53)	**45.87** (11.76)	112.15 (30.36)	118.11 (22.50)
S21	39.63 (12.36)	32.82 (11.36)**	134.46 (29.29)	**132.41** (24.09)
S23	50.20 (10.70)	48.72 (7.83)	135.96 (28.50)	**130.50** (23.18)
S24	43.74 (12.58)	40.20 (12.25)	122.19 (23.76)	**114.00** (21.89)
S25	44.42 (6.89)	**45.71** (6.54)	138.94 (16.60)	151.65 (21.10)***
S28	43.19 (12.91)	**48.60** (10.45)**	119.80 (21.77)	130.17 (19.45)*
S30	35.55 (8.95)	32.63 (6.56)	123.16 (21.81)	**113.20** (20.87)**
S32	47.78 (11.77)	43.61 (10.81)	125.35 (15.56)	**122.78** (14.47)
S35	44.43 (8.08)	34.94 (8.06)***	149.94 (20.29)	**143.29** (25.22)
S39	50.00 (12.25)	40.75 (9.21)***	139.60 (17.34)	141.99 (20.53)
S40	36.39 (11.31)	**38.03** (10.51)	119.76 (22.34)	128.26 (22.27)*
S42	54.20 (10.04)	51.94 (9.19)	109.65 (20.93)	118.34 (23.65)

VOT is for word-medial stops only. Features exhibiting divergence are bold and shaded. Features exhibiting hyperconvergence are in dashed boxes. Numbers in parentheses are standard deviations. Asterisks indicate significance levels of t-tests across conditions after Bonferroni correction. *$p < 0.05$, **$p < 0.01$, ***$p < 0.005$.

Each feature except vowel duration showed significant divergence for at least one participant. All participants had shorter mean VOT and lower mean vowel onset F0 in the pre-exposure condition than the model talker. One participant had a longer mean vowel duration than the model talker prior to exposure (S15); four participants had longer mean closure durations than the model talker prior to exposure (S23, S25, S35, S39).

Table 4.3 shows individual changes in VOT for cues that are present unambiguous in all positions in the word, as data from stop-initial consonants were incorporated into analysis on vowel duration and vowel F0. Given that the consonantal cues (closure duration, CSR, and overall stop duration) were only measured for word-medial consonants, Table 4.4 shows VOT changes just for consonants in these contexts.

Figure 4.4: Antagonistic accommodation (S25) and hyperconvergence (S35) in closure duration



S25 diverged in closure duration from the model ($p < 0.001$) and S35 hyperconverged to the model ($p < 0.001$). Lines show best-fit linear models for stop closure duration in baseline (Condition 1) and post-exposure (Condition 2) word-medial stops.

All but three speakers (S17, S24, S42) converged significantly in at least one feature. Two speakers (S25 and S42) diverged significantly in a feature — closure duration and onset vowel F0, respectively. Three speakers converged significantly in closure duration to the point of hyperconvergence; of these three, one had a shorter mean baseline closure duration than the model talker (S28), and two had longer mean durations (S35, S39). One example each of antagonistic accommodation and hyperconvergence in stop closure duration are illustrated in Figure 4.4. Only one speaker (S25) showed significant adjustment of overall stop duration across conditions.

After recording the stimuli, the model talker indicated that she had guessed the experiment had to do with the presence or absence of initial schwa in the various words. Intons-Peterson (1983) indicates that the effect of presented stimuli can be affected by the expectations of the experimenter, or in this case interlocutor. The model talker's reported focus on the context before the target consonants may be a contributing factor to the significance of Stop Position or Multisyllabicity as a predictor for each of the cues investigated.

4.3.4 Results by cue

VOT: As expected, most participants had a longer mean VOT in the post-exposure condition, with six participants showing a statistically significant increase ($p < 0.05$). In absolute terms, two participants diverged from the model VOT across all words (S20 and S28), while a third diverged specifically in the word-medial context (S30). This differing behavior depending on the environment of the stop, while not approaching the level of significance, indicates that a blanket assertion of convergence in VOT may be an oversimplification.

A mixed effects regression model predicting VOT was fit with Block (either *pre-exposure* or *post-exposure*), Vowel Identity, Stop Identity, Stop Position within the word, and following Vowel Onset F0 as predictor variables. Speaker and Word were entered as random effects. By-speaker random slopes of Block made significant contributions to model likelihood (see Table 4.5). The resulting model had an adjusted R^2 of 0.5523. Within-block Trial Position and following Vowel Duration did not prove to be significant predictors, nor did interactions between Stop Identity and Stop Position or between Block and Trial Position. Vowel F0 being a significant predictor is likely due to a bias in the placement of the boundary between consonant and vowel; dropping Vowel F0 as a predictor gave the model an adjusted R^2 of 0.5509.

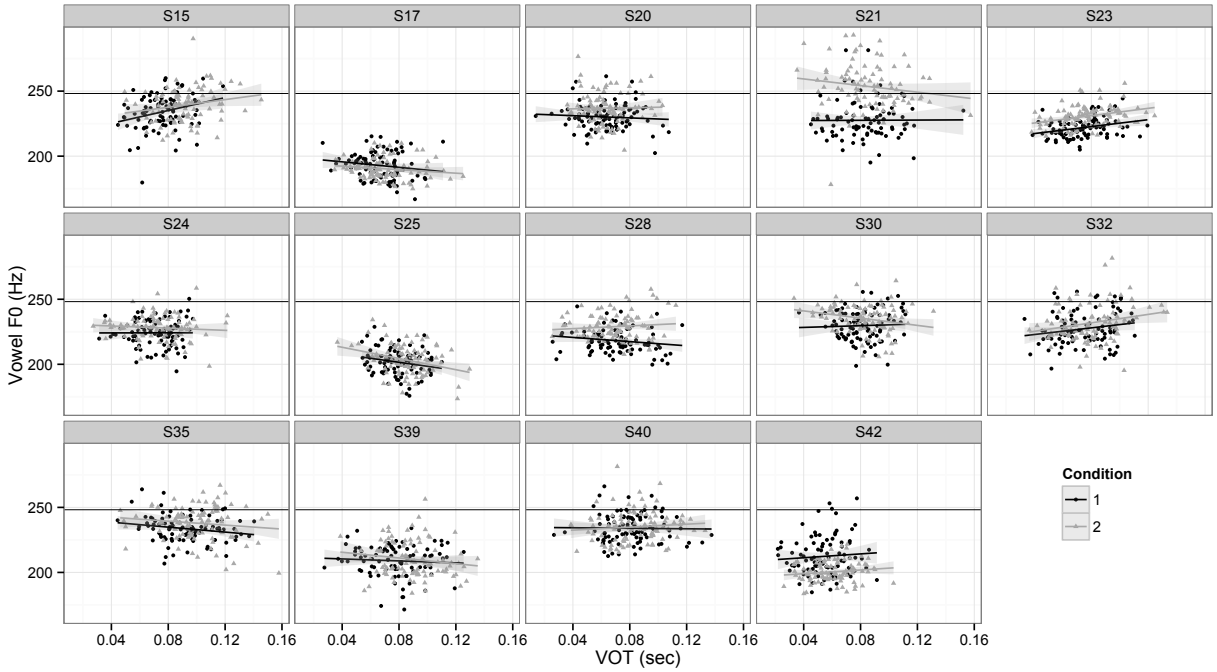
Table 4.5: Random slopes by speaker for VOT across conditions

	Slope (ms)		Slope (ms)
S15	***12.511	S28	0.256
S17	2.778	S30	3.172
S20	0.429	S32	5.241
S21	5.105	S35	*8.492
S23	4.399	S39	***10.048
S24	2.283	S40	3.670
S25	5.501	S42	3.620

Vowel onset F0: Overall, twelve of the fourteen participants converged toward the model's average onset F0, which was higher than that of all participants in the baseline condition. Two participants (S17 and S42) diverged in absolute average F0, including the single participant with the lowest baseline F0 (S17). One participant hyperconverged in F0 (S21). Best-fit lines for F0 as a function of VOT are shown in Figure 4.5 for each participant by condition.

A mixed effects regression model predicting vowel onset F0 was fit with Block, Vowel Identity, Stop Position within the word, and VOT duration as predictor variables. VOT being a significant predictor indicates that adjustment of F0 was automatic, contrary to prediction. Speaker and Word were entered as random effects. By-speaker random slopes of Block and Trial Position made significant contributions to model likelihood ($R^2 = 0.5572$). Stop Identity and Vowel Duration were not significant predictors, nor was Trial Position on its own. The interaction between Stop Position and Stop Identity also did not prove significant. Multisyllabicity (whether the word is monosyllabic or multisyllabic) was a worse predictor than Stop Position as a whole, and so was not included.

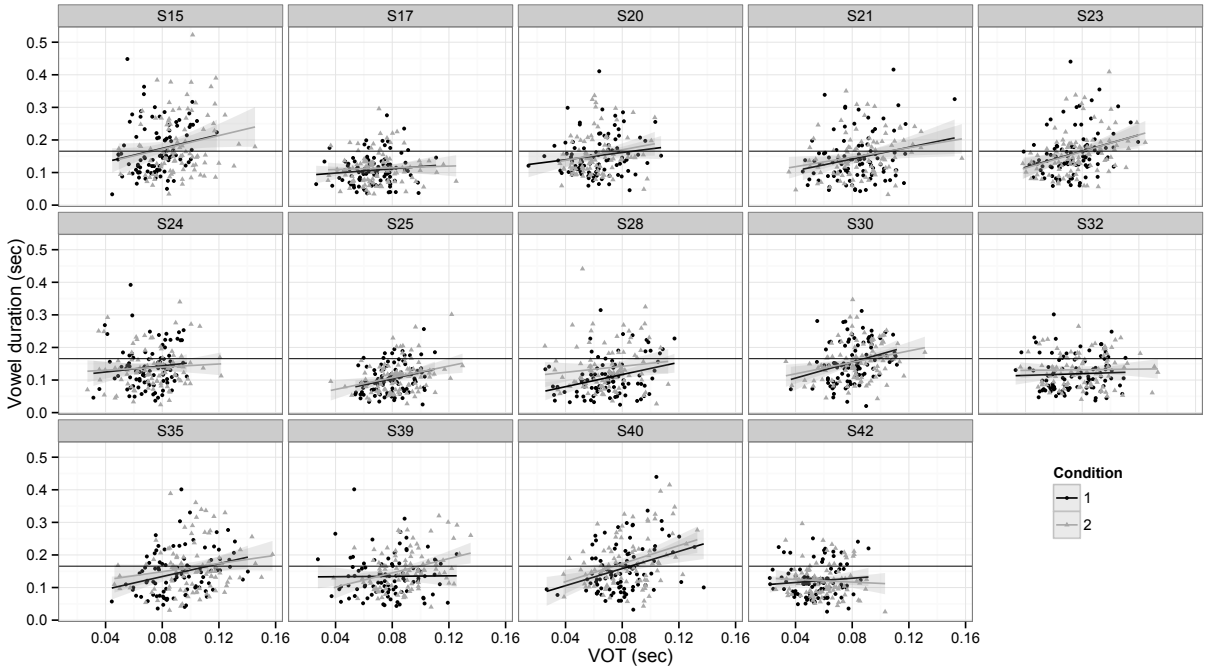
Figure 4.5: Vowel F0 by speaker



Condition 1 = pre-exposure, Condition 2 = post-exposure.
Model average onset vowel F0 of 248.18 Hz is shown for comparison.

Vowel duration: Overall, twelve of the fourteen participants converged toward the model's average onset vowel duration. Two participants (S15 and S24) diverged in absolute average vowel duration. One participant hyperconverged in vowel duration (S40). Best-fit lines for vowel duration as a function of VOT are shown for each participant by condition in Figure 4.6.

Figure 4.6: Vowel duration by speaker

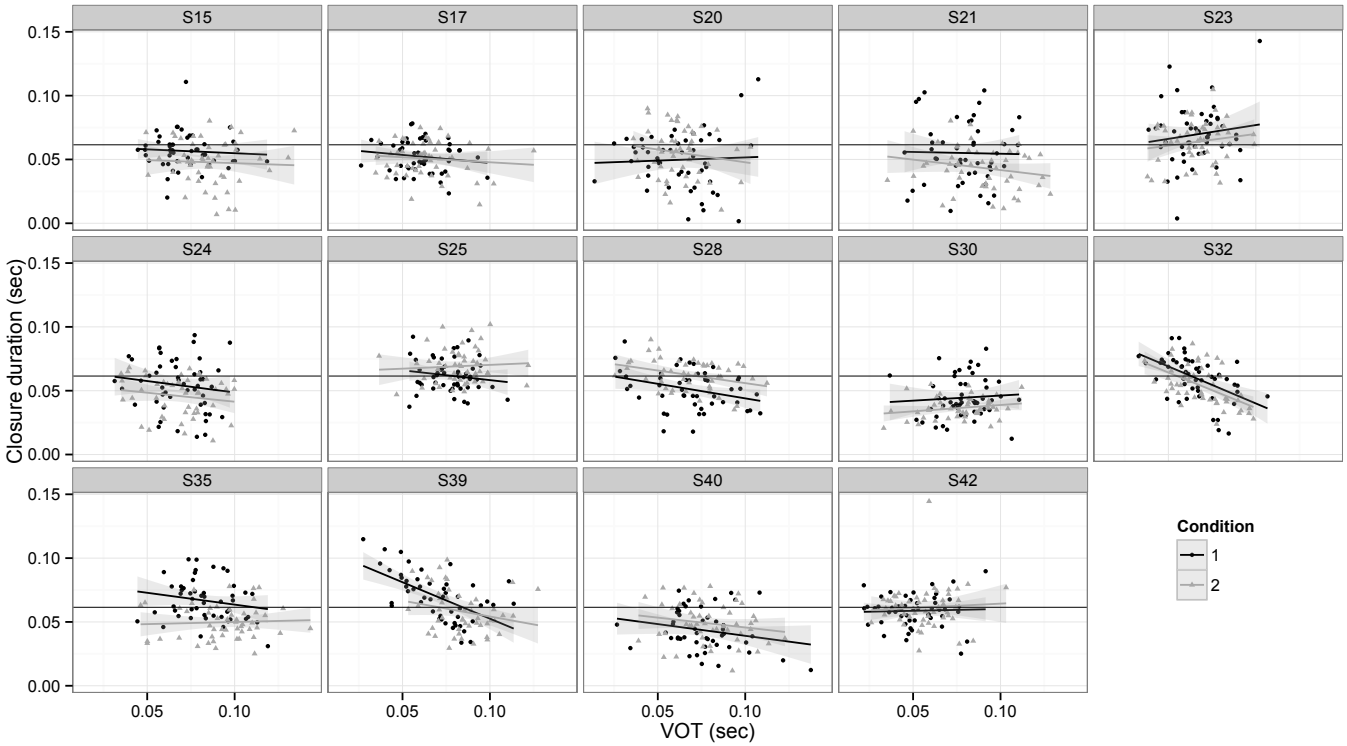


Condition 1 = pre-exposure, Condition 2 = post-exposure.
Model average vowel duration of 165.57 ms is shown for comparison.

A mixed effects regression model predicting vowel duration was fit with Block, Vowel Identity, Following Segment Type (voiced stop, voiceless stop, voiced fricative, liquid, nasal, vowel, or no following segment), Multisyllabicity, and within-block Trial Position as predictor variables. Speaker and Word were entered as random effects. By-speaker random slopes of Block and Trial Position made significant contributions to model likelihood, although Trial Position lost significance after Bonferroni correction (R^2 = 0.8197 with by-speaker Block). Stop position within the word and the identity of the following consonant did not prove to be significant predictors. Despite its strong correlation with vowel duration, VOT also was not a significant predictor for vowel duration, either by itself or in interaction with task condition. Since all but one speaker had pre-exposure vowels that were shorter than the model talker's vowels on average, this may indicate that speakers were accommodating to vowel duration independently from accommodation to VOT. While Multisyllabicity was a significant predictor of vowel duration, an interaction between Multisyllabicity and Block was not significant, indicating that changes to vowel duration were independent of the model talker's perceived speech rate.

Stop closure duration: Best-fit lines for closure duration as a function of VOT are shown for each participant by condition in Figure 4.7. Speakers with both longer (S25) and shorter (S15, S21, S24, S30) average closure durations than the model diverged in closure duration. This result indicates that divergence is not due to an intrinsic bias toward longer or shorter closure durations in themselves; moreover, it strongly suggests that this divergence is a side effect of convergence in other features.

Figure 4.7: Stop closure duration by speaker



Condition 1 = pre-exposure, Condition 2 = post-exposure.
Model average stop closure duration of 61.48 ms is shown for comparison.

A mixed effects regression model predicting stop closure duration was fit with Block, Trial Position, Stop Position within the word, Stop Identity, and VOT duration as predictor variables. Speaker and Word were entered as random effects. By-speaker random slopes of Block and Trial Position made significant contributions to model likelihood (R^2 = 0.4771). Across speakers, the model found global divergence in closure duration from the model talker of 2.64 milliseconds on average (β = –2.641 ms, standard error = 0.7768, t = –3.399).

4.4 Discussion

In accommodating to a model talker with artificially lengthened VOT, speakers were expected in the aggregate to converge toward model VOT, and in so doing diverge from the model talker in stop closure duration. Predictions for the other targeted coincident cues of VOT were contingent on individual speakers' values for those cues. Coincident cue adjustment was entertained but not expected for onset vowel F0, as the relationship between that cue and VOT was hypothesized to be a weak one. Following vowel duration was expected to undergo coincident cue adjustment, although not all speakers would have vowels that were shorter or longer than the model.

General findings: Of the three coincident cues to VOT under investigation — stop closure duration, onset vowel F0, and vowel duration — only the first two are significantly predicted by VOT across conditions. All three, however, change significantly after exposure to an external speech signal with extended VOT. This indicates that speakers accommodated independently to both VOT and vowel duration. Best models for vowel F0 and closure duration include Trial Position effects by speaker. This behavior is predicted by exemplar theories of phonetic organization (Johnson 1997, Pierrehumbert 2001). Best-fit models for vowel duration and closure duration also include Trial Position effects as fixed effects; while it is possible that these effects are those predicted by exemplar theories, the fact that these models were not improved with by-speaker effects suggests that they may be instead attributable to fatigue or neutralization over the course of the experiment.

Vowel F0 was significantly predicted by VOT across conditions. This coincident adjustment runs contrary to the hypothesis that vowel F0 would be controlled and adjusted independently of VOT. Also contrary to expectation, vowel duration was not significantly predicted by VOT across speakers, although half of speakers exhibited significant convergence to model vowel duration.

Closure duration is an interesting case. Exactly half of speakers exhibited absolute convergence in average closure duration, although only three did so with any statistical significance. Convergence occurred for speakers with mean pre-exposure closure durations that were both shorter than the model talker (S20, S28, S40, and S42) and longer than the model talker (S23, S35, S39). Likewise, divergence occurred for speakers with both shorter (S15, S17, S21, S24, S30, S31) and longer (S25) pre-exposure closure durations than the model talker. The fact that speakers diverged from the model closure duration both by lengthening and shortening their own closures indicates that this divergence is an instance of antagonistic accommodation.

Table 4.6: Correlations of characteristics and changes across conditions

	ΔVOT	ΔF0	ΔVowelDur	ΔClosureDur	ΔCSR	ΔStopDur
Power	0.321	−0.348	−0.577	0.155	0.138	−0.111
Extraversion	0.190	−0.034	−0.365	0.398	0.206	0.055
Agreeableness	0.239	−0.040	−0.109	0.456	0.252	−0.003
Conscientiousness	0.198	−0.563	−0.088	−0.055	0.136	−0.137
Neuroticism	0.228	−0.414	−0.573	−0.034	0.093	0.247
Openness	0.082	−0.399	−0.068	0.256	−0.038	0.137

Individual differences: Table 4.6 is a correlation matrix of the changes to measured acoustic features measured across conditions and the results of the personality and power questionnaires (Rogers et al. 1997; Saucier 1994) administered before the experiment. No correlation reached $p < 0.10$ after Bonferroni correction. Given the relatively small number of speakers analyzed, this lack of significance is not surprising.

However, it can be asserted that different participants behaved differently. Of the six speakers with a significant change to VOT and/or closure duration across conditions, two converged in both VOT and closure duration (S35, S39), two converged in VOT without adjusting closure duration (S15, S21), and two adjusted closure duration without adjusting VOT (S25, S28).

While S25 had an absolute average closure duration that was longer than that of the model talker, a two-tailed t-test did not show a significant difference in closure duration (t = –1.0054, df = 73.357, p = 0.318). Empirically, both S25 and S28 lengthened their closure duration; it cannot be stated with assurance whether they both converged in this regard.

Automaticity of coincident cues: Varying accounts of phonetic organization make differing predictions regarding how automatic the adjustment of coincident cues is. Kingston & Diehl's (1994) *phonetic reorganization* account holds that speakers are able to control cues of phonemic categories independently, as demonstrated by their ability to vary the phonetic realization of speech sounds between contexts. Phonetic implementations are "capacity-limited, attention-demanding, relatively easily learned and modified, and often accessible to conscious inspection" (Kingston & Diehl 1994). The phonetic reorganization account states that while some coincident cues are automatic due to physiological constraints on speakers' articulation, most coincident cues are controllable, including many of those that are phonologized in a language. This account predicts that controllable cues will not be automatically adjusted in concert with VOT; they will instead be independently adjusted in relation to the received speech signal's explicit value for those cues. In the context of this study, the phonetic reorganization account predicts that vowel F0 and vowel duration will not be automatically adjusted, whereas closure duration will. Although vowel duration was indeed adjusted independently of VOT, only two speakers (S20 and S23) showed independent adjustment of vowel F0. In the event that there is a physiological link between F0 and VOT, these findings are largely consistent with a phonetic reorganization account.

A contrary set of predictions is made by an *exemplar theory* account of phonetic organization (Johnson 1997; Pierrehumbert 2001). Exemplar theory holds that different subsets of experience are called on in deciding which variant of a given linguistic structure to use. Under this account, all coincident cues to VOT in English should be adjusted in accommodation to a signal with lengthened VOT, since all coincident cues are by definition associated in the main with VOT lengthening. This adjustment should take place even if the model for accommodation falls outside of the set of all prior experiences for the speaker, assuming that the speech signal is still analyzed as belonging to the same set of experiences. The most robust cues of VOT should evince the greatest degree of accommodation, due to speakers' increased familiarity with those cues (everybody uses them). These predictions appear to be borne out to some extent in the results of this study, as closure duration and vowel F0 were both adjusted in step with VOT. The lack of coincident adjustment of vowel duration may be attributed to the relatively large variation evinced by the model (sd = 76.5 ms).

Consonantal vs. vocalic accommodation: As has been alluded to, onset vowel F0 and vowel duration might properly be considered cues to the vowel following the manipulated consonant, rather than direct cues to the consonant itself. Only half of the speakers evinced significant levels of accommodation to both consonantal and vocalic cues. Most previous accommodation studies have restricted their area of inquiry to one or the other category: consonantal cues are generally easier to quantize, whereas vocalic cues are perhaps easier to perceive. The results of this experiment indicate that the conflation of these two classes of targets for accommodation may lead to erroneous conclusions about how much accommodation takes place.

How to measure accommodation: As Pardo (2013) notes in her assessment of the state of the art in phonetic convergence, the generality of convergence phenomena actually poses a

problem for researchers trying to measure convergence absolutely: "it is not known which acoustic attributes are perceptible to listeners, and which play a relatively minor role. On the one hand, a unit change in intensity is not perceived in the same [manner] as a unit change in F0/pitch. On the other hand, convergence in one acoustic attribute might offset divergence in another" (2013:3).

In any study investigating the effects of sociological factors such as interlocutor attractiveness or likeability, it would not be enough to measure phonetic convergence simply in terms of VOT. As these findings indicate, participants S20 and S28 would not be seen as converging in such a study. However, it is not at all clear that they diverged, or even that they failed to show convergence: both speakers converged in vowel duration, and S28 also converged in vowel F0 and in closure duration. Given that speakers generally seem not to be able to independently manipulate VOT and closure duration at the same time, it is misleading to expect convergence along every feature — and in a study examining only one feature, that feature may well be the "wrong one" for some participants. Of course, this understanding leads to the question: is total divergence even possible? If enough features are measured, participants will inevitably exhibit convergence in at least one. For a speaker in this study to exhibit total divergence, they would have to either shorten their mean VOT and lengthen an already-long closure duration, or have an uncommonly long VOT pre-exposure. Nielsen (2011) found that speakers would not shorten VOT in voiceless stops in accommodation to a model signal with shortened VOT. She suggested that this might be due to the possibility of introducing phonological ambiguity, given that voiced and voiceless stops in English are differentiated primarily by the comparatively longer VOT in voiceless stops. This finding indicates that there are circumstances in which total divergence will not happen. But this does not mean that quantitative measurement of accommodation is straightforward.

On the other hand, qualitative measurement of accommodation is no more straightforward. AXB similarity tasks used to qualitatively confirm accommodation (Goldinger 1998, Nye & Fowler 2003) will not disambiguate between the adjustment of different cues to accommodation. It is likely that some cues are more salient in some fashion to speakers than others. However, the comparative salience of a feature may not be universal across a population. Given the robustness of VOT's accommodation effects and its crucial role in English phonology, it is sensible to expect VOT to be one of these salient cues. In that light is surprising that as many speakers displayed significant adjustment to closure duration as did to VOT. It may be possible to look at whether speakers who do not accommodate to VOT evince this lack of salience in their production of stops — for example, they may invariably voice phonologically voiced stops whereas other American English speakers do not.

Ramifications for sound change: If convergence is the default state of affairs in interlocution, antagonistic accommodation may be an inevitable byproduct thereof. As such, it is conceivable to delineate a course for sound change in which no misperception is necessary on the part of the listener. Antagonistic accommodation is analogous to Ohala's formulation of hypercorrection (1989, 1993), in that it is an "inappropriate application of [...] corrective rules" (Ohala 1989): speakers making attendant adjustments to their production of a received target. The key difference is that the 'correcting' is not due to a mismatch between a speaker's intended production and a listener's perception of that intention, rather between a speaker's intended production and their ability to effect that production.

Chapter 5: Conclusion

My goal in this dissertation has been to examine patterns of linguistic behavior that are convergent to the speech of another, and to investigate the idea that these patterns have a uniform motivation. Overall, there are four main points that I intend to highlight.

5.1 Convergence behavior is difficult — but possible — to isolate and measure

There are so many potential sources of variation in speech behavior for a given individual that it is difficult to attribute observed variation to a particular source. Accurately measuring convergence requires correctly identifying the source of variation as stemming from responses to received stimuli, rather than from internal factors (see also Pardo 2013). Additionally, once an individual has been exposed to a stimulus that is repeated, it is difficult to attribute subsequent changes in behavior to a particular token of that stimulus, meaning that it is far from straightforward to measure changes in convergence behavior over time. In measuring change, it is necessary to contrast change over a short period of time with change over a longer period of time. There is a further difficulty in that a given individual may converge to different linguistic features at different rates[25]. Several of the experiments discussed in this dissertation each grapple with aspects of these inherent difficulties.

In Chapter 2 I reported the results of a Mechanical Turk shadowing experiment looking at the effects of word frequency and power on convergence to vowel formants. I exposed participants to high-frequency, low-frequency, and no-frequency (pseudo-)words, expecting differential degrees of convergence between these three groups. Participants heard each pseudoword only once over the course of the experiment, in order to maintain their no-frequency status. Unfortunately, this may have limited the robustness of statistical analysis for the results of this experiment: participants did not evince significant differences in the rate of convergence to pseudowords as distinct from either high- or low-frequency words.

In Chapter 3, I discussed an experiment designed to compare convergence across different domains of linguistic structure. In order to increase the likelihood that changes in individuals' speech were attributable to convergence, I used a dyadic game task in participants made repeated utterances within a prescribed purposive context. The game's two players made requests in turn for objects with multiple common names in English, allowing me to analyze lexical convergence (which name was used for a given object) and syntactic convergence (whether a request was in the syntactic form of a question or a statement), as well as phonetic convergence (in vowel formants and vowel duration), within the same interaction.

Chapter 4 pursued the ramifications of varying rates of convergence to different linguistic features. In the experiment discussed therein, experiment participants in a laboratory setting heard speech with manipulated VOT in English voiceless stops, but natural values for related features, including the closure duration of those stops, and the onset f0 and duration of the following vowel. Participants converging towards the model talker in one dimension simultaneously diverged in one or more related dimensions, indicating the difficulty of isolating a particular feature or set of features as the sole locus of imitation. This means that a study intended to examine convergence

[25] See also Section 6.2.

behavior that analyzes a single feature may well be analyzing the "wrong feature" for some participants.

Chapter 4 also raised the idea that speakers may evince different convergence patterns to vowels as opposed to consonants. Only half of the participants in the study in Chapter 4 evinced significant levels of accommodation to both consonantal cues (VOT; stop closure duration) and vocalic cues (onset vowel F0; vowel duration) within the same context.

5.2 Convergence behavior is not specific to particular domains of linguistic structure

If perceived stimuli are stored as episodes, and production is informed by the recall of those episodes, then convergence behavior is consequent. If input is stored multiply corresponding to multiple domains of linguistic structure, and those domains are recalled independently, then we expect convergence behavior in each of those domains, although it will be realized differentially within each domain depending on the constitution of that domain's relevant episodes.

Within this dissertation I am assuming this multiple storage of episodes, and I am referring to each domain of (linguistic) structure within which episodes may be stored as a *categorization schema*. According to the approach advocated here, episodes are only stored within currently active categorization schemata, and each episode is stored in all active categories whose definitions it is perceived to match. In this approach, convergence is not a discrete process, and it may indeed not be confined to language.

In Chapter 1, I discussed two general patterns of convergence. A given experience may be stored as both an episode of *what is being done* and as an episode of *the way something is being done*, depending on the definition and purpose of the categorization schema in question. Convergence toward what is being done is a "higher-level", goal-oriented type of convergence; convergence toward the way something is being done is a "lower-level", procedure-oriented type of convergence.

The dyadic game experiment discussed in Chapter 3 was designed to allow analysis of phonetic, lexical, and syntactic convergence within the same interaction. Across participants, I found no correlation between convergence behavior at any of these levels of linguistic structure. This lack of correlation is contrary to the predictions of Interactive Alignment Theory (Pickering and Garrod 2004), in which alignment at one level of linguistic representation leads to alignment at other levels due to interconnections between the levels. According to the account I discuss in this dissertation, it is not the case that different levels of representation influence each other in this regard; rather, multiple levels of representation are likely to be activated at the same time due to commonalities in the situational context that make those multiple levels germane concurrently.

5.3 Convergence behavior is predicted by personal autonomy

Differences in convergence behavior between individuals may be attributable, in whole or in part, to differences in individuals' social characteristics. Power is a potentially relevant social characteristic, as it is associated with increased attention to stereotypes (Goodwin et al. 2000) and reflexive social cognition (Keltner et al. 2003). Additionally, previous studies have found a relationship between convergence rates and participants' conversational role, which may be attributable to their relative power in an interaction (Pardo et al. 2010).

Several of the experiments discussed within this dissertation took participants' self-evaluated power into account. In the shadowing experiment detailed in Chapter 2, I looked at personal power as a potential predictor of convergence to the vowel formants of a model talker, but found no effect. Surprisingly, however, the results of the dyadic game experiment detailed in Chapter 3 indicated that individuals with higher self-ratings for *community activism and autonomy* converged less than those with lower autonomy self-ratings, across domains of linguistic structure. The factor for autonomy approached or reached significance as a predictor of convergence to features at each of phonetic (vowel formants), lexical, syntactic, and utterance levels of linguistic representation. This pattern suggests an inverse relationship between a person's sense of autonomy and their attention to the low-level details of others' behavior.

5.4 Convergence behavior is a potential actuator of sound change

In Chapter 4, I discussed the potential ramifications of convergence for theories of sound change. While previous researchers have pointed out the potential for convergence to be a mechanism for the transmission of a sound change (Delvaux & Soquet 2007), the experiment detailed in Chapter 4 provided evidence that convergence can result in the appearance of new variants within an interaction. In this study I looked at *non-faithful transference* of phonetic features during accommodation to the speech of another. Study participants converging towards a model talker in one dimension simultaneously diverged in a related dimension. Specifically, while participants evinced overall convergence toward the VOT of an English-speaking model talker's voiceless stops, they globally diverged from the model talker's closure duration for those stops. This *antagonistic* relationship between the two features led to the possibility of the creation of a new variant, in the form of a new set of corresponding values for those features.

Antagonistic accommodation then provides a possible course for the actuation of sound change (Weinreich et al. 1968). Moreover, it provides a course that does not require a mismatch between a speaker's intended production and a listener's perception of that intention (cf. Ohala 1989, 1993). Instead, the locus of change lies in the difference between a speaker's intended production and their ability to effect that production. However, this potential mismatch hinges on the gradience of phonetic features. As such, antagonistic accommodation cannot account for the actuation of diachronic changes to categorical features.

References:

Abrego-Collier, C., Grove, J., Sonderegger, M., & Yu, A. C. 2011. Effects of speaker evaluation on phonetic convergence. In *Proceedings of the 17th International Congress of the Phonetic Sciences:* 192–195.

Aguilar, L. J. 2011. *Gender identity threat in same & mixed-gender negotiations: Speech accommodation & relational outcomes*. Doctoral dissertation.

Aguilar, L., Downey, G., Krauss, R., Pardo, J., Lane, S., & Bolger, N. 2015. A Dyadic Perspective on Speech Accommodation and Social Connection: Both Partners' Rejection Sensitivity Matters. *Journal of Personality*. doi:10.1111/jopy.12149

Ambady, N., & Rosenthal, R. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 111(2): 256.

Anderson, C., John, O.P., & Keltner, D. 2012. The Personal Sense of Power. *Journal of Personality* 80(2). doi: 10.1111/j.1467-6494.2011.00734.x

Aylett, M., & Turk, A. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America* 119(5): 3048–3058.

Babel, M. E. 2009. *Phonetic and social selectivity in speech accommodation*. Doctoral dissertation.

Babel, M. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39(04): 437–456.

Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40(1): 177–189.

Babel, M. & Bulatov, D. 2012. The role of fundamental frequency in phonetic accommodation. *Language and Speech* 552:231–248.

Babel, M., McGuire, G., Walters, S., and Nicholls, A. 2014. Novelty and social preference in phonetic accommodation. *Journal of Laboratory Phonology* 5(1): 123–150. doi:10.1515/lp-2014-0006

Bates, D., Maechler, M., Bolker, B., and Walker, S. 2016. "lme4: Linear mixed-effects models using Eigen and S4." R package.

Bekkering, H., Wohlschläger, A., and Gattis, M. 2000. Imitation of Gestures in Children is Goal-directed. *The Quarterly Journal of Experimental Psychology* 53A (1): 153–164.

Bock, J. K. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3): 355–387.

Bock, J. K., Dell, G. S., Chang, F., & Onishi, K. H. 2007. Structural persistence from language comprehension to language production. *Cognition* 104: 437–458.

Bock, J. K., & Griffin, Z. M. 2000. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129: 177–192.

Boersma, P. & Weenink, D. 2014. Praat: doing phonetics by computer. Version 5.3.62, retrieved 2 January 2014 from http://www.praat.org/.

Boucher, V. J. 2002. Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics* 64(1): 121–130.

Bourhis, R. Y. & Giles, H. 1977. The Language of Intergroup Distinctiveness. In *Language, Ethnicity, and Intergroup Relations*, ed. H. Giles. 119–135.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. 2000. Syntactic co-ordination in dialogue. *Cognition*, *75*(2): B13–B25.

Brennan, S.E. & Clark, H.H. 1996. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Phonology: Learning, Memory, & Cognition* 22(6): 1482–1493.

Brouwer, S., Mitterer, H., & Huettig, F. 2010. Shadowing reduced speech and alignment. *The Journal of the Acoustical Society of America* 128(1): EL32–37.

Brysbaert, M., & New, B. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4): 977–990.

Buhrmester, M., Kwang, T., & Gosling, S. D. 2011. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6(1): 3–5.

Chang, C. B. 2012. Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40(2): 249–268.

Charpentier, F., and Stella, M. 1986. Diphone synthesis using an overlap-add technique for speech waveforms concatenation. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86* 11: 2015–2018.

Chistovich, L., Fant, G., de Serpa-Leitao, A., & Tjernlund, P. 1966. Mimicking of synthetic vowels. *Quarterly Progress and Status Report, Speech Transmission Lab, Royal Institute of Technology, Stockholm* 1(2): 1–18.

Clark, H.H. & Wilkes-Gibbes, D. (1986). Referring as a collaborative process. *Cognition* 22: 1–39.

Costa, P. T., Jr. & McCrae, R. R. 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Odessa, FL: Psychological Assessment Resources.

Crowne, D. P. & Marlowe, D. 1960. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 24(4): 349–354.

Delvaux, V. & Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64(2-3): 145–173.

Dias, J. W. & Rosenblum, L. D. 2011. Visual influences on interactive speech alignment. *Perception-London* 40(12): 1457–1466.

Dijksterhuis, A. & Bargh, J. 2001. The Perception-Behavior Expressway: Automatic Effects of Social Perception on Social Behavior. *Advances in experimental social psychology* (33): 1–40.

Downey, G. & Feldman, S. I. 1996. Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology* 70 (6): 1327–1343.

Engle, R. W. 2002. Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science* 11(1): 19–23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. A. 1999. Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach. *Journal of Experimental Psychology* 128(3): 309–331.

Ferreira, V. S., Bock, K., Wilson, M. P., and Cohen, N. J. 2008. Memory for Syntax Despite Amnesia. *Psychological Science* 19(9): 940–946.

Fiebach, C. J., Friederici, A. D., Müller, K., and von Cramon, D. Y. fMRI Evidence for Dual Routes to the Mental Lexicon in Visual Word Recognition. *Journal of Cognitive Neuroscience* 14(1): 11–23.

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

Garrod, S. & Anderson, A. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27: 181–218.

Garrod, S. & Clark, A. 1993. The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes* 8(1): 101–126.

Gentilucci, M. & Bernardis, P. 2007. Imitation during phoneme production. *Neuropsychologia* 45(3): 608–615.

Giles, H. 1971. *A study of speech patterns in social interaction: Accent evaluation and accent change*. Unpublished Ph.D thesis, University of Bristol.

Giles, H., Taylor, D. M., & Bourhis, R. 1973. Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society* 2(2): 177–192.

Giles, H., Coupland, J., & Coupland, N. 1991. Accommodation theory: Communication, context, and consequence. In *Contexts of accommodation: Developments in applied sociolinguistics*: 1–68.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2): 251–279.

Goldinger, S. D. 2013. The cognitive basis of spontaneous imitation: Evidence from the visual world. *Proceedings of Meetings on Acoustics* 19: 060136. doi: 10.1121/1.4800039.

Goldinger, S. D. & Azuma, T. 2004. Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review* 11(4): 716–722.

Gregory Jr, S. W. 1983. A quantitative analysis of temporal symmetry in microsocial relations. *American Sociological Review 1983*: 129–135.

Gregory, S. W. 1990. Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior* 14: 237–251.

Gregory Jr, S. W. & Hoyt, B. R. 1982. Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research* 11(1): 35–46.

Gries, S. Th. 2005. Syntactic priming: A corpus-based approach. *Journal of Psychological Research* 34(4): 365–399.

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. 2008. Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, *58*(2): 214–238.

Holt, L. L., Lotto, A. J., & Kluender, K. R. 2001. Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?. *The Journal of the Acoustical Society of America* 109(2): 764–774.

Holtgraves, T. 1986. Language structure in social interaction: Perceptions of direct and indirect speech acts and interactants who use them. *Journal of Personality and Social Psychology* 51(2): 305–314.

Hombert, J. M. 1976. The effect of aspiration on the fundamental frequency of the following vowel. *Proceedings of the 2nd Annual Meeting of the Berkeley Linguistics Society*, 212–219.

Honorof, D. N., Weihing, J., & Fowler, C. A. 2011. Articulatory events are imitated under rapid shadowing. *Journal of Phonetics* 391(1): 18–38.

House, A. S., & Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25(1): 105–113.
Intons-Peterson, M. J. 1983. Imagery paradigms: How vulnerable are they to experimenters' expectations? *Journal of Experimental Psychology: Human Perception and Performance* 9(3): 394–412.
Jaffe, J., & Feldstein, S. 1970. *Rhythms of dialogue*. New York: Academic Press.
Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. *Talker Variability in Speech Processing*: 145–165.
Keltner, D., Gruenfeld, D. H., & Anderson, C. 2003. Power, approach, and inhibition. *Psychological Review* 110(2): 265–284.
Kent, R. 1979. Imitation of synthesized English and non-English vowels by children and adults. *Journal of Psycholinguistic Research* 8(1): 43–60.
Kim, M., Horton, W. S., & Bradlow, A. R. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology* 21:125–156.
Kingston, J., & Diehl, R. L. 1994. Phonetic knowledge. *Language* 70(3): 419–454.
Kootstra, G. J., et al. 2010. Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language* 63(2): 210–231. doi:10.1016/j.jml.2010.03.006
Kraljic, T., Brennan, S. E., & Samuel, A. G. 2008. Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107(1): 54–81.
Kuhl, P. K. & Meltzoff, A. N. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America* 100(4):2425–2438.
Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. 2016. lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0–32. https://CRAN.R-project.org/package=lmerTest
Labov, W. 2001. Principles of linguistic change, Volume 2: Social factors. Oxford: Blackwell.
Lehiste, I. 1970. *Suprasegmentals*. Cambridge: Cambridge University Press.
Levitan, R. & Hirschberg, J. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Interspeech* 2011: 3081–3084.
Lewandowski, N. 2012. Automaticity and consciousness in phonetic convergence. *The Listening Talker*: 71. Conference proceedings.
Lisker, L. 1986. "Voicing" in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29(1): 3–11.
Miller, R. M., Sanchez, K., & Rosenblum, L. D. 2013. Is speech alignment to talkers or tasks? *Attention, Perception, & Psychophysics* 75(8): 1817–1826.
Mitterer, H. & Ernestus, M. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109(1): 168–173.
Mixdorff, H., Cole, J., & Shattuck-Hufnagel, S. 2012. Prosodic Similarity: Evidence from an Imitation Study. In *Speech Prosody 2012*.
Namy, L. L., Nygaard, L. C., & Sauerteig, D. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21(4): 422–432.

Natale, M. 1975a. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32(5): 790–804.
Natale, M. 1975b. Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills* 40(3): 827–830.
Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1): 62–85.
Nielsen, K. 2008. *The specificity of allophonic variability and its implications for accounts of speech perception*. Los Angeles: University of California, Los Angeles dissertation.
Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2): 132–142.
Nielsen, K. 2014. Phonetic Imitation by Young Children and Its Developmental Changes. *Journal of Speech, Language, and Hearing Research* 576: 2065–2075.
Nye, P. W. & Fowler, C. A. 2003. Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics* 31(1): 63–79.
Ohala, J. J. 1989. Sound change is drawn from a pool of synchronic variation. *Language change: Contributions to the study of its causes*, 173–198.
Ohala, J. J. 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and Perspectives*. 237–278.
Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119(4): 2382–2393.
Pardo, J. S. 2013. Measuring phonetic convergence in speech production. *Frontiers in Psychology* 4(559). doi: 10.3389/fpsyg.2013.00559
Pardo, J. S., Jay, I. C., & Krauss, R. M. 2010. Conversational role influences speech imitation. *Attention, Perception, & Psychophysics* 72(8): 2254–2264.
Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 401:190–197.
Pardo, J. S., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. 2013. Influence of role-switching on phonetic convergence in conversation. *Discourse Processes* 504:276–300.
Pickering, M. J. & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(2): 169–189.
Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper, eds., *Frequency and the Emergence of Linguistic Structure*: 137–157.
R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rogers, E. S., Chamberlin, J., Ellison, M. L., & Crean, T. 1997. A consumer-constructed scale to measure empowerment among users of mental health services. *Psychiatric Services* 48(8): 1042–1047.
Sapir, E. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Co.
Schneider, W., & Shiffrin, R. M. 1977. Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention. *Psychological Review* 84(1): 1–66.
Shockley, K., Sabadini, L., and Fowler, C. A. 2004. Imitation in shadowing words. *Perception & Psychophysics* 66(3): 422–429.

Snodgrass, J. G. & Vanderwart, M. 1980. A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *Journal of Experimental Psychology: Human Learning and Memory* 6(2): 174–215.

Strand, E. A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1): 86–99.

Strand, E. A. and Johnson, K. 1996. "Gradient and Visual Speaker Normalization in the Perception of Fricatives." In *KONVENS*, pp. 14–26.

Street Jr, R. L. & Cappella, J. N. 1989. Social and linguistic factors influencing adaptation in children's speech. *Journal of Psycholinguistic Research* 18(5): 497–519.

Sui, J. & Liu, C. H. 2009. Can beauty be ignored? Effects of facial attractiveness on covert attention. *Psychonomic Bulletin & Review* 16: 276. doi:10.3758/PBR.16.2.276

Sumner, M. & Samuel, A. G. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60(4): 487–501.

Ullman, M. T. 2001. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30(1): 37–69.

Ullman, M. T. 2004. Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92: 231–270.

Vaan, L., Schreuder, R. and Baayen, R. H. 2007. Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon* 2: 1-23.

Walters, S. A., Babel, M. E., & McGuire, G. 2013. The role of voice similarity in accommodation. *Proceedings of Meetings on Acoustics* 19(1): 060047.

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. 2014. Socially-mediated syntactic alignment. *Language Variation and Change* 26(3): 387–420.

Webb, J. T. 1970. Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. In A. W. Siegman & B. Pope, eds., *Studies in Dyadic Communication*: 55–70.

Weinreich, U., Labov, W., & Herzog, M. I. 1968. *Empirical foundations for a theory of language change*. University of Texas Press.

Yu, A. C. L. 2010. Perceptual compensation is correlated with individuals' "autistic" traits: implications for models of sound change. *PloS one* 5(8): e11950.

Yu, A. C. L. 2013. Individual differences in socio-cognitive processing and sound change. *Origins of Sound Patterns: Approaches to phonologization*. Oxford University Press.

Yuan, J. and Liberman, M. 2008. Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08*.

Zellou, G., Scarborough, R., & Nielsen, K. 2013. Imitability of contextual vowel nasalization and interactions with lexical neighborhood density. *Proceedings of Meetings on Acoustics* 19(1): 060083.

Appendix A: Table of stimulus word frequency (Chapter 2).

Word	Freq	LgFreq	Word	Freq	LgFreq	Word	Freq	LgFreq
badly	26.20	3.1261	*julep*	0.35	1.2788	*signal*	37.76	3.2849
beauty	48.24	3.3911	*kiosk*	0.45	1.3802	*snicker*	0.29	1.2041
bebop	0.41	1.3424	*knowledge*	25.53	3.1149	*social*	33.39	3.2315
bonsai	0.98	1.7076	kosher	2.69	2.1399	*soldier*	38.92	3.2980
boring	27.41	3.1458	*lockjaw*	0.27	1.1761	*spigot*	0.29	1.2041
bottle	50.75	3.4131	marriage	77.06	3.5945	*spoken*	19.84	3.0056
burlap	0.22	1.0792	*meaning*	37.33	3.2799	*squeegee*	0.27	1.1761
buses	3.94	2.3054	*mission*	47.06	3.3804	*stagnant*	0.27	1.1761
butane	0.29	1.2041	*modern*	18.24	2.9689	standard	18.43	2.9736
cackle	0.29	1.2041	*Mohawk*	0.47	1.3979	*stogie*	0.35	1.2788
cherry	13.59	2.8414	*muumuu*	0.29	1.2041	*stolen*	34.31	3.2433
chicken	61.73	3.4982	*ouzo*	0.35	1.2788	*student*	43.04	3.3416
china	24.86	3.1035	*pallet*	0.41	1.3424	sugar	37.75	3.2849
closely	9.18	2.6712	panther	2.57	2.1206	*tadpole*	0.59	1.4914
common	44.61	3.3572	paper	103.35	3.7220	*talent*	26.12	3.1248
contact	64.80	3.5193	parent	13.14	2.8267	target	37.96	3.2871
contour	0.27	1.1761	*pathos*	0.24	1.1139	*teacher*	55.73	3.4538
currents	1.69	1.9395	perfect	158.65	3.9081	tension	8.55	2.6405
cushion	2.16	2.0453	*phallic*	0.29	1.2041	tested	10.53	2.7308
custom	6.20	2.5011	*pigment*	0.29	1.2041	*theory*	28.61	3.1644
dachshund	0.35	1.2788	*pivot*	0.45	1.3802	*ticket*	45.57	3.3664
decent	28.10	3.1565	*pizza*	33.51	3.2330	tiger	18.53	2.9759
dirty	66.45	3.5302	*plastic*	18.76	2.9814	tizzy	0.18	1.0000
earache	0.29	1.2041	*pocket*	35.71	3.2605	*tonsil*	0.35	1.2788
easel	0.27	1.1761	*poultice*	0.29	1.2041	tourney	0.14	0.9031
facet	0.25	1.1461	*practice*	45.69	3.3675	*travel*	33.37	3.2312
female	31.61	3.2076	*program*	42.63	3.3375	*Tuesday*	23.65	3.0817
fifty	18.82	2.9827	*proper*	25.27	3.1106	tundra	0.27	1.1761
final	49.67	3.4038	*pseudo*	0.29	1.2041	twisted	10.50	2.7332
foamy	0.27	1.1761	punish	9.67	2.6937	*unit*	36.18	3.2662
follow	123.20	3.7982	*quotient*	0.33	1.2533	*value*	21.51	3.0406
foolish	17.51	2.9513	*reflux*	0.35	1.2788	*victim*	47.73	3.3865
forage	0.31	1.2304	rider	7.71	2.5955	*village*	33.57	3.2338
fritter	0.29	1.2041	sample	14.59	2.8722	weapon	46.65	3.3766
frosty	2.37	2.0864	*season*	31.47	3.2057	welfare	7.88	2.6053
gather	15.67	2.9031	*seesaw*	0.43	1.3617	*whisker*	0.25	1.1461
grimace	0.33	1.2553	shimmered	0.04	0.4771	winding	1.96	2.0043
helix	0.39	1.3222	*shoulder*	26.20	3.1261	*witness*	51.39	3.4186
hobble	0.35	1.2788	*shoplift*	0.33	1.2553	*yokel*	0.31	1.2304
jacket	33.41	3.2317	sickness	7.94	2.6085	zebra	2.51	2.1106
journey	19.94	3.0077						

Words in plaintext only appeared in the baseline condition (in both blocks). Words in italics appeared in both the baseline and stimulus conditions. Underlined words only appeared in the stimulus condition. Word frequencies are from the SUBTLEX-US corpus.

Appendix B: Nonce words used in frequency task (Chapter 2).

backom	*coldick*	*gillard*	*peamack*	*steener*
/ˈbæ.kəm/	/ˈkol.dɪk/	/ˈgɪl.ɚd/	/ˈpiː.mæk/	/ˈstiː.nɚ/
beeda	*contiff*	*hottice*	*podger*	*stiggan*
/ˈbiː.də/	/ˈkɑn.tɪf/	/ˈhɑ.rɪs/	/ˈpɑː.dʒɚ/	/ˈstɪ.gən/
beudon	*dassome*	*keachous*	*poonid*	*stotion*
/ˈbjuː.dən/	/ˈdæ.səm/	/ˈkiː.tʃəs/	/ˈpuː.nɪd/	/ˈstoʊ.ʃən/
bickman	*deason*	*lockage*	*porrid*	*taston*
/ˈbɪk.mən/	/ˈdiː.zən/	/ˈlɑː.kədʒ/	/ˈpo.ɹɪd/	/ˈtæ.stən/
boadgie	*dision*	*moalen*	*seachal*	*teefle*
/ˈboʊ.dʒi/	/ˈdɪ.ʒən/	/ˈmoʊ.lən/	/ˈsiː.tʃəl/	/ˈtiː.fəl/
bocko	*doolick*	*nobbit*	*shostum*	*togrous*
/ˈbɑ.koʊ/	/ˈduː.lɪk/	/ˈnɑ.bɪt/	/ˈʃɑ.stəm/	/ˈtoʊ.gɹəs/
calit	*fallor*	*pactor*	*soakle*	*tuker*
/ˈkæ.lɪt/	/ˈfæ.lɚ/	/ˈpæk.tɚ/	/ˈsoʊ.kəl/	/ˈtuː.kɚ/
chicket	*fignous*	*pathent*	*soozle*	*witsick*
/ˈtʃɪ.kət/	/ˈfɪg.nəs/	/ˈpæ.θənt/	/ˈsuː.zəl/	/ˈwɪt.sək/

Appendix D: Baseline word list for game experiment (Chapter 3).

had
head
hid
hide
hood
who'd
heed
hod
hawed
Hud

badly
beauty
buses
canoe
chicken
china
closely
cushion
dirty
fifty
follow
frosty
perfect
response
rider
shimmered
sickness
sugar
twisted
winding

Appendix E: Table of stimulus word frequency (Chapter 4).

Word	Freq	LgFreq	Word	Freq	LgFreq	Word	Freq	LgFreq
a cappella	0.43	1.3617	cord	7.02	2.5551	pend†		
accompany	4.75	2.3856	cost	54.92	3.4475	petunia	2.08	2.0294
accord	1.63	1.9243	count	89.96	3.6617	picante	0.04	0.4771
accost	0.06	0.6021	cues	0.69	1.5563	picard	1.53	1.8976
account	44.71	3.3581	cult	4.45	2.3579	pinion	0.16	0.9542
accuse	5.69	2.4639	curd	0.43	1.3617	point	236.53	4.0815
accustom	0.12	0.8451	currents	1.69	1.9395	posable	0.02	0.301
acquaint	0.39	1.3222	custom	6.20	2.5011	potato	11.29	2.7612
acquire	2.65	2.1335	cute	87.75	3.6509	potential	18.82	2.9827
acquit	0.47	1.3979	department	63.84	3.5128	quaint	2.18	2.0492
acute	2.94	2.179	detention	6.53	2.5237	quit	90.10	3.6624
akin	0.27	1.1761	earache	0.29	1.2041	recognize	34.31	3.2433
all	5161.86	5.4204	else	449.16	4.36	repent	2.41	2.0934
apace	0.06	0.6021	interest	50.94	3.4148	retaining	0.65	1.5315
apart	47.02	3.38	katana	0.08	0.699	risk	49.04	3.3983
apiece	3.96	2.3075	kin	4.27	2.3404	satirical	0.18	1
appall	0.04	0.4771	kosher	2.69	2.1399	standard	18.43	2.9736
apparent	4.22	2.3345	look	1947.27	4.997	tack	2.12	2.0374
appeal	13.00	2.8222	macaw	0.24	1.1139	tall	32.33	3.2175
appearing	2.33	2.0792	marriage	77.06	3.5945	target	37.96	3.2871
appease	0.49	1.415	memory	48.57	3.3941	taskmaster	0.16	0.9542
append†			natural	42.35	3.3347	tasty	6.31	2.5092
appoint	2.04	2.0212	nose	69.75	3.5512	tempt	2.53	2.1139
Atari†			occult	1.57	1.9085	tend	12.27	2.7973
atone	0.84	1.6435	occurred	14.45	2.8681	tension	8.55	2.6405
attack	75.55	3.5859	occurrence	1.18	1.7853	tested	10.53	2.7308
attempt	19.12	2.9894	opinion	42.00	3.331	tiger	18.53	2.9759
attend	14.02	2.8549	opposable	0.35	1.2788	tire	12.37	2.8007
attention	98.67	3.7018	pace	9.57	2.6893	tizzy	0.18	1
attested	0.12	0.8451	pall	0.45	1.3802	tofu	2.69	2.1399
attire	1.49	1.8865	panther	2.57	2.1206	tomato	5.90	2.48
attorney	40.39	3.3141	paper	103.35	3.722	tone	16.86	2.935
attune	0.02	0.301	parent	13.14	2.8267	torque	0.73	1.5798
battalion	5.84	2.4757	part	261.51	4.1251	tourney	0.14	0.9031
botanical	1.04	1.7324	pastry*	1.92	1.9956	tundra	0.27	1.1761
Capone	1.63	1.9243	pawn	4.33	2.3464	tune	15.61	2.9015
catastrophe	2.47	2.1038	peace	69.61	3.5504	upon	62.73	3.5051
catatonic	0.78	1.6128	peas	4.65	2.3766	welfare	7.88	2.6053
choir	5.31	2.4346	peel	5.35	2.4378	wish	235.12	4.0789
company	147.20	3.8755	peering	0.33	1.2553	yes	1996.76	5.0079
Copernicus	0.67	1.5441						

†Did not occur in the production list. *Does not occur in the SUBTLEX-US corpus.