Top-Down Effects on Speech Perception:

An Integrated Computational and Behavioral Approach

by

Neal P. Fox

B. A., University of Virginia, 2009

Sc. M., Brown University, 2012

Submitted in partial fulfillment of the requirements

for the Degree of Doctor of Philosophy in the

Department of Cognitive, Linguistic, and Psychological Sciences at Brown University

Providence, Rhode Island

May 2016

This dissertation by Neal P. Fox is accepted in its present form

by the Department of Cognitive, Linguistic, and Psychological Sciences

as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

_____	_____
Date	Sheila E. Blumstein, Advisor

Recommended to the Graduate Council

_____	_____
Date	Michael J. Frank, Reader
_____	_____
Date	James L. Morgan, Reader

Approved by the Graduate Council

_____	_____
Date	Peter M. Weber, Dean of the Graduate School

Neal P. Fox

Education

Brown University	Ph.D. in Cognitive Science	January 2016
	Top-Down Effects on Speech Perception: An Integrated Computational and Behavioral Approach (Advisor: Dr. Sheila E. Blumstein)	
Brown University	M.S. in Cognitive Science	May 2012
	Top-down effects from syntactic category expectations on speech processing	
University of Virginia	B.A. with Distinction in Cognitive Science	May 2009
	Minor: Mathematics; Post-grad. training in Systems Engineering (2009-10)	

Honors and Awards

Reisman Brain Science Graduate Fellowship	Brown Institute for Brain Science	2014
Dissertation Fellowship	Brown University	2013
Best Poster Award	Society for Teaching of Psychology	2013
NSF Graduate Research Fellowship	National Science Foundation	2010–2013
LSA Summer Institute Fellowship	Linguistic Society of America	2011
Calvin & Rose G Hoffman Prize	The Marlowe Society	2011
Graduate Fellowship	Brown University	2010
Raven Society (Academic Honor Society)	University of Virginia	2010
Graduate Research Fellowship in Engineering	University of Virginia	2009–2010

Publications

1. **Fox, N. P.,** Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language, 83*, 97-117.
2. **Fox, N. P.** & Blumstein, S. E. (in press). Top-down effects of syntactic sentential context on phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*.
3. Luthra, S., **Fox, N. P.**, & Blumstein, S. E. (2016). Speaker information affects false memory recognition of unstudied lexical-semantic associates. (under revision)
4. Caplan, S., **Fox, N. P.**, McClosky, D. M., & Charniak, E. (2016). Lexical substitution for cross-domain parser adaptation. (under revision)
5. Reilly, M., Guediche, S., **Fox, N. P.**, & Blumstein, S. E. (2016, manuscript). Articulatory planning and motor reprogramming: An fMRI investigation. (in prep)
6. **Fox, N. P.**, Blumstein, S. E., & Frank, M. J. (2016, manuscript). Bayesian Integration of Acoustic and Sentential Evidence in Speech: The BIASES Model of Spoken Word Recognition in Context.
7. **Fox, N. P.** & Blumstein, S. E. (2016, manuscript). Bottom-up and top-down contributions to lexical processing deficits in aphasia.

Refereed Conference Presentations

1. **Fox, N. P.** & Larsen, E. W. (2009). A comparative optimality theoretic outlook on loanword phonology in Huave dialects of Mexico. Rice University Linguistics Society Third Biennial Meeting; Houston, TX.
2. **Fox, N. P.** (2012). Top-down effect of syntactic category expectations on spoken word recognition. CUNY Conference on Sentence Processing; New York, NY.
3. **Fox, N. P.**, Ehmoda, O., & Charniak, E. (2012). Statistical stylometrics and the Marlowe-Shakespeare authorship debate. Georgetown University Roundtable on Languages and Linguistics; Washington DC.
4. **Fox, N. P.** & Blumstein, S. E. (2013). Top-down effects from sentence context on speech processing in aphasia. Society for Neurobiology of Language; San Diego, CA.

5. **Fox, N. P.** & Reilly, M. (2013). Significantly different: A meta-analysis of the gap between statistics students learn and statistics psychologists use. Northeast Conference for Teachers of Psychology; Bridgeport, CT. *(Best Poster Award)*
6. **Fox, N. P.**, Reilly, M., & Blumstein, S. E. (2014). Independent and interacting effects of sentential context and phonological neighborhood structure in spoken word production. Acoustical Society of America; Providence, RI.
7. Luthra, S., **Fox, N. P.** & Blumstein, S. E. (2015). Speaker information affects false recognition of unstudied lexical-semantic associates. Society for Neurobiology of Language; Chicago, IL.
8. **Fox, N. P.** & Blumstein, S. E. (2015). Computational and neural mechanisms of top-down effects on speech perception. Society for Neurobiology of Language; Chicago, IL.

Teaching Experience

Introduction to Cognitive Neuroscience	Teaching Assistant	Brown University	2013
Quantitative Methods in Psychology	Teaching Assistant	Brown University	2012
Computational Cognitive Science	Teaching Assistant	Brown University	2012

Teaching Development and Leadership

Teaching Certifications:

Certificate III: Professional Development Seminar	2014
Certificate IV: The Teaching Consultant	2013
Certificate I: Reflective Teaching Seminar	2012

Program Facilitation and Teaching Consultation:

Certificate I: Reflective Teaching Seminar	2012–2015
Center for Engaged Learning Seminar: Undergraduate Research Mentorship	2014
Principles & Practice in Reflective Mentorship	2013–2014
Senior Teaching Consultant	2013–2014
New Teaching Assistant Orientation: Interactive Classrooms	2013

Undergraduate Research Mentorship

Spencer Caplan, Brown class of 2015	2013–2015
Sahil Luthra, Brown class of 2014	2012–2014

Service

Session Chair, Speech Perception, Acoustical Society of America Meeting	2014
CLPS Department Representative; Sheridan Center for Teaching & Learning	2012–2014
Graduate Representative, Campus Access Advisory Committee	2011–2014
Session Chair, Computational Linguistics, Georgetown Roundtable on Lang. & Ling.	2012
Graduate Representative, Brown Univ. Presidential Search Advisory Committee	2011–2012

Professional Memberships

Society for Neurobiology of Language	2013–2016
Acoustical Society of America	2013–2016
American Psychological Association	2013–2016
Society for the Teaching of Psychology	2013–2016
Northeast Psychological Association	2013–2016
Linguistic Society of America	2011–2015
Institute for Electrical and Electronic Engineers	2009–2010
IEEE Engineering in Medicine and Biology Society	2009–2010
Society for Neuroscience	2008–2010

Research Interests

computational models of speech and language perception and production, cognitive neuroscience of language and aphasia, computational linguistics and natural language processing

Birth Information: Neal P. Fox was born on August 3, 1989, in San Diego, CA.

Acknowledgements

"It is clear that the successful completion of graduate school requires the recruitment of resources from a variety of sources." Eye rolls aside, this statement is undoubtedly one of the most fundamental truths of graduate school. I have been lucky to work with, learn from, and get support from incredible scientists and outstanding people.

At every turn, Sheila Blumstein was there to guide me, to back me up, and – when necessary – to kick me in the butt. The role of advisor and mentor is often a thankless one, but it is impossible to express the impact Sheila has had on me and on my career. The breadth and depth of her work will always be an inspiration to me, along with the earnestness and fierceness of her loyalty, to say nothing of her unparalleled patience. I was also fortunate to work with Michael J. Frank. Michael's multidisciplinary approach to cognitive science and computational neuroscience and his willingness to explore new questions opened many doors to me. Throughout my time at Brown, Jim Morgan was a valuable sounding board for my scientific work, but he is also representative of why the CLPS Department has become as much a home as a workplace. It was a privilege to "grow up" in a department in which a full professor would invite a graduate student to his house on a Saturday morning so the student can use his power tools to complete construction of a pair of Cornhole boards. Eugene Charniak's mentorship and collaboration provided me with an indispensible perspective at every step of my graduate career. I was often impressed his humility when he didn't know the answer (which was rare) and his ability to discern worthwhile projects from interesting exercises (while still seeing the value in both).

The mentorship of several other faculty members and postdocs in the CLPS department has also been critical to my development as a scientist. Especially at the beginning, Laura Kertz, Paul Allopenna, Emily Myers, Rachel Theodore and Hugh Rabagliati were longsuffering teachers and mentors. I am also grateful to Thomas Serre, Kathy Spoehr and David Badre for their formative mentorship in my three teaching assistantships in the department. Finally, I would be remiss if I failed to thank the dedicated folks who were there for me (and the entire department) whenever I asked for help and worked behind the scenes every day so I wouldn't need to. I feel lucky to call Reinette, Don, Jesse, Michelle, Rosa and Bill my friends.

My research projects spanned three labs (Sheila's, Michael's and Eugene's) over the last several years, and members of each lab have helped me clarify, shape, and also *do* much of that research. Among those colleagues were Megan Reilly, Sahil Luthra, Sara Guediche, John Mertus, Kathy Kurowski, Jeff Cockburn, Nick Franklin, Anne Franklin, Matt Nassar, Jason Scimeca, Micha Elsner, Spencer Caplan, Dave McClosky, Rebecca Mason and Omran Ehmoda. Additionally, several funding sources were critical to my research and training, including a Graduate Research Fellowship (DGE 0228243) from the National Science Foundation, a Graduate Fellowship from the Brown Institute for Brain Science, a summer training fellowship from the Linguistic Society of America, and funding from Brown's Graduate School and the CLPS Department.

As much as graduate school is and should be a *disciplinary* endeavor, I cannot overstate the amount of personal and professional development I gained by participating in the Sheridan Center's programming, and especially thanks to the incredible mentorship of the Center's former director Kathy Takayama. Working with her and the rest of the

staff and students associated with the Sheridan Center was unquestionably one of the most important opportunities of my graduate career. Similarly, the opportunities to meet new colleagues, mentors, and friends from across the school by serving on University committees were formative experiences and fulfilling outlets during my time at Brown.

All work and no play would have made grad school a bummer. Luckily, I met Megan Reilly on my first day of my first year at Brown, so there was never any chance of that happening. She and the rest of the DuckTales Cohort (*awoohoo*: Jason Scimeca, Chris Erb, Patrick Heck, David Mély) made me a better scientist and person, and they put up with me when I constantly pointed out top-down effects on speech perception in real life (TDEB). Without them, these last few years would have been DISGUSTING.

Finally, as families go, I've got the best one. Thanks, Mom, Dad, Sean, Collin and Alyson. And thanks to the newest members of my family, Emily and Ginny. You have been the sources of many, many smiles while I wrote this thesis, and I will never be able to re-bay you.

Table of Contents

Chapter 3

Exploring and Evaluating the BIASES Model of Spoken Word Recognition in Context

---

Chapter 4

Top-Down Effects on Spoken Word Recognition in Aphasia:
A Model-Based Assessment of Information Processing Impairments

List of Tables

List of Figures

List of Figures (continued)

Introduction

During auditory language comprehension, a listener's principal objective is to infer what meaning the speaker intended to convey. For a healthy adult communicating in his or her native language, this task is typically both effortless and errorless. What makes this behavioral generalization noteworthy is the fact that there is rarely just one possible interpretation of a given acoustic signal; in fact, perceptual uncertainty is ubiquitous in speech communication. Indeed, listeners are faced with uncertainties arising from countless sources ranging from unclearly produced speech to imperfect listening conditions to inescapable ambiguities inherent in language (e.g., homophony). It is easy to see how the challenge posed by such pervasive uncertainty might be crippling to a speech processing system that relied exclusively on these ambiguous acoustic cues available in the perceived speech signal to decode a speaker's meaning. A fundamental question, then, regards how the perception of speech can be so robust despite these barriers.

In the present work, it is argued that at least part of the answer to that question is that even though much ambiguity exists, when one source of information is unreliable, there are usually other cues available in the signal that can be leveraged to understand the speaker's intended meaning. For instance, although the spoken word /*bɔrd*/ could be an exemplar of either the word *board* or of *bored*, words are rarely uttered in isolation, and – most of the time – there is little doubt as to which meaning to assign to the lexically ambiguous speech token. This is, in large part, because the word's linguistic and extra-linguistic context provides another cue that can aid in the extraction of meaning, particularly when the acoustic cues are degraded or insufficient.

While this explanation may appear straightforward, it raises the key question of how multiple sources of information are integrated by the speech processing system. It is that question which is the focus of this thesis. More specifically, the present work employs behavioral experiments and computational modeling in order to investigate how so-called *bottom-up* acoustic cues available in the sensory signal and *top-down* information about which words or sounds are likely in a given context are integrated during online auditory language comprehension.

For the purpose of exposition, let a word be defined as the discrete linguistic unit that stands at the juncture between the sound information perceived by the listener in the speech signal and the underlying meaning associated with the signal (Blumstein, 2009). Spoken word recognition, then, represents a critical sub-routine of auditory language comprehension if a listener is to extract meaning from the perceived signal. A given word can be thought of as being associated with (1) a lexical form that defines how the word sounds, and (2) the word's meaning. When a speech token that resembles the lexical form of a word is perceived, that word can be recognized and its meaning accessed.

When it comes to recognizing a spoken word, acoustic cues in the sensory signal are certainly the paramount source of information available to the speech perception system. The perceptual system is adept at decoding the speech signal based on auditory cues alone. These information sources, which are the product of low-level sensory and phonetic processing of the input, are referred to as bottom-up cues.

However, as important as bottom-up cues are, much research has shown that, in addition to integrating a host of bottom-up cues, word recognition also involves the recruitment of top-down information that is not immediately available in the signal, but

instead relies on cognitive or higher-level linguistic processing. A general conclusion of this line of research is that listeners tend to perceive things that are more probable; for instance, identification is biased towards words rather than non-words (Ganong, 1980) and towards contextually consistent or sensible words over words that are inconsistent or nonsensical given the context (e.g., Borsky et al, 1998; Fox & Blumstein, in press).

Although the roles of both bottom-up and top-down cues are well attested, this thesis examines the basic computational principles that underlie their integration. Although a number of models have sought to tackle the question of how top-down cues come to influence speech perception, several issues remain unclear. Firstly, a long-standing, much-debated topic regards whether the observed biases in listeners' responses (toward words over non-words and toward contextually consistent words over inconsistent words) reflect the direct, top-down modulation of perceptual processing of the input or whether they reflect processing biases at a later decision-making level (see, e.g., McClelland, Mirman & Holt, 2006; Norris, McQueen & Cutler, 2000). Secondly, despite substantial evidence that sentential context influences spoken word recognition, existing models lack an explicit characterization that can account for these effects. Thirdly, another weakness of existing spoken word recognition models is that they largely ignore the enormous variability that exists in the observed sizes of top-down effects. Finally, the extent to which patients with aphasia experience deficits in top-down processing and cue integration during speech perception is poorly understood.

In the present work, each of these four issues is considered in turn. Empirical and computational methodologies are employed in order to probe the questions each issue poses. Ultimately, the results of this thesis provide a more complete picture of the

computations that take place at the interface between the perceptual processing of speech and the cognitive and linguistic processing of language, while also establishing a novel theoretical basis that promises to guide future work.

Chapter 1[1]

Top-down effects of syntactic sentential context on phonetic processing

1.1. Introduction

During auditory language comprehension, listeners integrate information from a variety of sources in their categorization of sounds and words, especially when confronted with degraded or ambiguous speech. Besides low-level acoustic cues, listeners are also sensitive to higher-level information that is not immediately available in the raw sensory input. For instance, listeners exhibit a lexical bias in their categorization of a phonetically ambiguous segment between */g/* and */k/* such that they label the segment as */g/* more often when followed by *–ift*, but as */k/* more often when followed by *–iss* (Ganong, 1980; see also Burton, Baum & Blumstein, 1989; Burton & Blumstein, 1995; Connine, 1990; Connine & Clifton, 1987; Fox, 1984; McQueen, 1991; Miller & Dexter, 1988; Myers & Blumstein, 2008; Pitt, 1995; Pitt & Samuel, 1993).

Moreover, when a stimulus is phonetically ambiguous between two words (e.g., between *goat* and *coat*), listeners exhibit a semantic bias such that they label the ambiguous word as *goat* more often when embedded in a sentence like *The busy farmer hurried to milk the…* but as *coat* more often in *The elderly tailor had to dry-clean the…* (Borsky, Tuller & Shapiro, 1998; Connine, 1987; Connine, Blasko & Hall, 1991; Garnes & Bond, 1976; Guediche, Salvata & Blumstein, 2013; Miller, Green & Schermer, 1984). Like semantic information, syntactic (Isenberg, Walker & Ryder, 1980; van Alphen & McQueen, 2001), morphosyntactic (Martin, Monahan & Samuel, 2012), and pragmatic

[1] At the time of submission of this dissertation, a version of Chapter 1 is currently in press at the *Journal of Experimental Psychology: Human Perception and Performance* (http://dx.doi.org/10.1037/a0039965). This article may not exactly replicate the

information (Do, 2011; Rohde & Ettlinger, 2012) have all been shown to bias listeners' identifications of phonetically ambiguous words.

From this robust literature, it is clear that sensory processing alone cannot explain listeners' judgments about the identities of spoken words and sounds. In order for higher-level information to influence spoken word recognition, perceptual input must make contact with lexical representations which act as a gateway to the semantic, syntactic and other properties of words, and those lexical representations must then be able to influence behavioral responses in tasks like those described above (Samuel, 2011).

1.1.1. Models of Spoken Word Recognition: Competing Frameworks

There exists, however, a longstanding debate about how those lexical representations come to influence spoken word recognition (for reviews, see McClelland, Mirman & Holt, 2006; McQueen, Norris & Cutler, 2006). Two competing families of spoken word recognition models – *interactive models* and *autonomous models* – each account for the effects of higher-level cues by appealing to different mechanisms. In particular, they differ in how pre-lexical (i.e., perceptual/phonetic) representations and lexical representations influence one another. Both approaches allow for a bottom-up flow of information such that pre-lexical processing of speech modulates the extent to which competing lexical representations are supported. However, only *interactive models* of spoken word recognition, exemplified by TRACE (McClelland & Elman, 1986; McClelland, 1991), incorporate top-down feedback projections that, conversely, allow lexical representations to modulate the extent to which competing pre-lexical representations are supported (see also Adaptive Resonance Theory; Grossberg, 1980, 2003; Grossberg & Myers, 2000).

Autonomous models of spoken word recognition, on the other hand, eschew top-down modulation of phonetic processing. Instead, they account for lexical and contextual effects on listeners' responses by positing that both higher-level and lower-level information can influence phonemic decisions, but it is maintained that pre-lexical representations remain faithful to the bottom-up acoustic input. Thus, under the autonomous view, the observed biases reflect the integration of multiple information sources, but, crucially, this integration does not affect lower-level phonetic processing itself (see, e.g., Norris, McQueen & Cutler, 2000; McQueen, Jesse & Norris, 2009). A succession of autonomous models has been proposed in the literature, including Race (Cutler & Norris, 1979; Cutler, Mehler, Norris & Segui, 1987), Shortlist (Norris, 1994), Merge (Norris et al., 2000), and Shortlist's Bayesian implementation (Norris & McQueen, 2008). Although each varies in its details, none allows for higher-level modulation of phonetic processing through feedback (McQueen et al., 2006; see also Fuzzy Logical Model of Speech Perception; Massaro, 1989).

Thus, any behavioral demonstration of lexical or contextual biases in phoneme judgments could, in theory, be explained at the level of the judgment itself (a *post-perceptual* explanation, as in autonomous models), or by direct modulation of pre-lexical processing prior to the judgment (a *perceptual* explanation, as in interactive models). Because of this, interactive and autonomous models of spoken word recognition have, in practice, proven difficult to distinguish. However, past work suggests that these two frameworks may diverge in their predictions about the time course of these effects.

1.1.2. Time Course of Sentential Context Effects

One result that proponents of autonomous models have cited as incompatible with TRACE (and interactive models more generally) concerns the time course of lexical and contextual effects. Specifically, they contend that if top-down feedback can directly alter the activation of pre-lexical representations (as it can in TRACE), then the influence of higher-level information could only grow or remain stable as a function of processing time, but could not diminish (McQueen, 1991; Tuinman, Mitterer & Cutler, 2014; van Alphen & McQueen, 2001). According to this argument, top-down biasing information within an interactive framework tends to overwrite the ambiguous bottom-up input pattern, shaping it and pulling it towards a pattern that would be expected for lexically- or contextually-consistent speech.

For instance, in one experiment, the identification of phonetically ambiguous function words (between *de* and *te*; roughly *the* and *to* in Dutch) was shown to be biased by manipulating the target words' grammaticality in context (van Alphen & McQueen, 2001). Ambiguous stimuli were labeled as /*de*/ more often in sentences like *We verstoppen* [***?***] *schaatsen* (*We hide* [*the*]/[*to*] *skates; de-biased*) than in sentences like *We behoren* [***?***] *schaatsen* (*We ought* [*to*]/[*the*] *skate; te-biased]* (*cf.* Isenberg, Walker & Ryder, 1980). Importantly, though, when responses were divided into bins based on reaction time (*cf.* Fox, 1984), contextual biases from an immediately preceding syntactic cue were strongest in fast responses and grew weaker with time. A similar pattern of results was found for the time course of syntactic context effects on a different type of phonetic judgment by Tuinman, Mitterer and Cutler (2014).

Can interactive models account for a smaller contextual bias in slower responses than in faster ones? Van Alphen and McQueen (2001) argue that they cannot: if

ambiguous bottom-up information has been overwritten by top-down feedback such that even early activation levels at the pre-lexical level are biased by context, then the original (ambiguous) pre-lexical representation cannot be recovered in order to yield more ambiguous (i.e., less biased) responses later in processing. That is, in interactive models, feedback permanently and irrevocably biases the bottom-up record of the unbiased pre-lexical representation (Massaro, 1989). Although Dahan, Magnuson and Tanenhaus (2001) dismiss this argument on the grounds that using response latencies to track top-down influences on pre-lexical activation "is not straightforward" (p. 321), it remains unclear to what extent the time course of sentential context effects on spoken word recognition does, in fact, challenge interactive models of speech perception.

The present work aimed to examine this question by testing two possible alternative explanations of previous time course data (van Alphen & McQueen, 2001; Tuinman et al, 2014). Specifically, two experiments investigated whether characteristics of the experimental designs in earlier work might have allowed subjects to adopt strategies that would not only explain the diminishing bias effect, but also undermine the ability of those data to distinguish between interactive and autonomous models. Following previous work that showed diminishing contextual biases, both of the present experiments examined syntactic sentence context effects on subjects' perception of phonetically ambiguous speech. Experiment 1.1 tested whether the diminishing influence of sentential context would persist when subjects could not plan contextually congruent responses prior to the presentation of the target stimulus. Experiment 1.2 tested whether the diminishing bias effect would persist when subjects were induced to engage in phoneme identification rather than word identification strategies.

1.2. Experiment 1.1

In prior work (van Alphen & McQueen, 2001; Tuinman et al, 2014), the experimental design allowed subjects to identify the contextually appropriate response before hearing the ambiguous target segment. For example, in van Alphen and McQueen's (2001) study, a preceding context biased the identification of an acoustically ambiguous target between *de* and *te*. Because this experiment utilized a single continuum with only two alternatives (*de* and *te*), subjects could have identified and prepared a grammatically congruent response before they encountered the target. This raises the question of whether some responses might reflect decisions that were generated before processing of the target could have actually begun. If the responses that were fastest were disproportionately contaminated with such pre-planned decisions, it would not be surprising to observe a fast-arising bias that appears to weaken at longer RTs (once subjects had heard the target word). Importantly, this explanation would not challenge interactive models; button-presses planned before a stimulus is presented could not bear on whether the pre-lexical representation of that stimulus is modulated by interactive feedback.

Thus, Experiment 1.1 utilized two voice-onset time (VOT) continua, rather than just one: a noun–verb continuum (*bay–pay*) and a verb–noun continuum (*buy–pie*), crossing target word voicing with syntactic category bias. Tokens from these VOT continua were appended to noun-biasing and verb-biasing sentence contexts (e.g., *Valerie hated the...*, *Brett hated to...*), and subjects identified the initial segment of sentence-final targets (/*b*/ vs. /*p*/). In this way, subjects could not know which phonemic response (/*b*/ vs. /*p*/) was congruent with the contextual bias of a sentence until the target word was

presented, thereby preventing them from anticipating or planning a specific button-press response before hearing the target word.

1.2.1. Methods

1.2.1.1. Materials

1.2.1.1.1. Target Word Selection

Sixteen monolingual volunteers who spoke American English participated in a norming study to confirm that the four critical targets words (*bay*, *pay*; *buy*, *pie*) had strong syntactic category biases in the expected directions. The four targets were included in a randomly ordered list of 40 words that included words from a variety of syntactic categories. For each word in the list, subjects wrote one sentence "that might be heard in everyday speech," and their responses were coded for the target words' part of speech usage in each sentence. The noun targets (*bay* and *pie*) were each used by at least 15 of 16 subjects as a noun; the verb targets (*pay* and *buy*) were each used by at least 14 subjects as a verb. Thus, *bay*/*pay* and *buy*/*pie* were judged to be sufficiently biased noun/verb and verb/noun minimal pairs, respectively.

1.2.1.1.2. Sentence Contexts

Twenty main verbs (e.g., *hate*, *want*) that could be followed by either a noun phrase or an infinitive phrase (e.g., *hate the bay; hate to pay*) were identified. Forty sentence contexts (20 noun-biased; 20 verb-biased) were then constructed by concatenating a first name, the past tense form of the main verb, and either *the* or *to*, yielding pairs of sentence contexts like *Valerie hated the...* and *Brett hated to...* The full list of contexts can be found in Appendix A. This design helped ensure that participants could not use information from the main verb to predict the target word. In this way, any

syntactic bias effect on responses could be attributed to the influence of the immediately preceding function word.

1.2.1.1.3. Stimulus Recording

Sentences ending with the target words were recorded by a female monolingual native American English speaker in a sound-dampened room with an Edirol digital recorder (model R09-HR; Sony microphone model ECM-MS907; sampled: 44,100 Hz / 24 bits / stereo; resampled in BLISS speech-editing software: 22,050 Hz / 16 bits / mono; Mertus, 1989). Target words were spliced out of the sentences' waveforms yielding 40 partial sentences (20 noun-biased, 20 verb-biased).

1.2.1.1.4. Target Word Manipulation

A natural token of *bay* and of *buy* served as base tokens for two VOT continua, constructed using the BLISS waveform editor (Mertus, 1989). Beginning with the base voiced token of *bay*, an acoustically modified voiceless end of the *bay–pay* continuum and each intermediary token were generated by successively adding aspiration from the middle of the aspiration of a naturally-produced *pay* token and removing pitch periods of equal duration from the onset of the vowel of the natural *bay* token. This procedure yielded 12 stimuli with VOTs ranging from 2 to 64 milliseconds. The onset's burst was amplified (2x) for all tokens in the continuum because the aspiration from the *pay* token rendered the burst in the natural *bay* inaudible. Tokens of the *buy–pie* continuum were generated in the same manner, yielding 12 stimuli with VOTs ranging from 3 to 62 milliseconds. As with the *bay–pay* continuum, the onset's burst was amplified (3x) for all tokens in the continuum. Twenty milliseconds of silence was appended to the beginning of each token of both continua. The waveforms of all stimuli across the two continua

were normalized for amplitude so that the highest peaks of the waveforms were equally high.

1.2.1.1.5. Target Word Token Selection

Ten monolingual native English-speaking volunteers from the Brown University community participated in a norming study whose goal was to select the eight target stimuli: two ambiguous tokens and one token from each phonetic category endpoint for each continuum. Twenty trials each of the 12 tokens of each continuum (480 total trials) were presented in isolation (without sentential context) binaurally to participants in random order. Participants responded whether each target began with a "p" or "b" by pressing a corresponding button (response mapping was counterbalanced between subjects) and were instructed to respond as quickly as possible while maintaining accuracy, and to guess if they were unsure.

The two tokens from each continuum with the identification rates closest to 50% and the highest mean response reaction times (Pisoni & Tash, 1974) were selected as the ambiguous tokens from each continuum. Endpoint tokens of each continuum were selected such that each the /*b*/ and /*p*/ was equidistant from the ambiguous pair. Table 1.1 shows the results for the eight selected tokens.

Continuum	Token #	VOT (ms)	Mean % p	Mean RT (ms)
bay–pay	2	7	1.5	591
bay–pay	4	18	42.5	794
bay–pay	5	24	74	778
bay–pay	7	35	97.5	655
buy–pie	2	7	2	609
buy–pie	4	18	39	852
buy–pie	5	23	79	826
buy–pie	7	34	86.5	720

Table 1.1. Mean classification rates and reaction times (RTs) for the selected tokens from each voice-onset time (VOT) continuum

1.2.1.2. Participants

Fifty self-reported native monolingual American English speakers with normal hearing from the Brown University community volunteered or received course credit to participate in Experiment 1.1. None had participated in any of the norming studies reported earlier. Due to technical difficulties, one subject's incomplete data were excluded from analysis.

1.2.1.3. Task

Each of the eight selected tokens was appended to each of the 40 sentence contexts, yielding 320 stimuli. The resulting design crossed two levels of CONTINUUM (*bay–pay*, *buy–pie*) with two levels of CONTEXT (noun-biased, verb-biased) and four tokens from each VOT continuum. All sentences were presented binaurally in a random order after eight practice trials. Participants were instructed to indicate whether the last word in each sentence began with a "b" or a "p" by pushing the appropriate button with either the index or middle finger of their dominant hand (response mapping was counterbalanced between subjects). The experiment did not advance to the next trial until a subject responded, but participants were instructed to respond as quickly as possible

while maintaining accuracy, and to guess if they did not know. They were also warned that some sentences might not make sense. Reaction times were measured from the onset of the target word.

1.2.2. Results

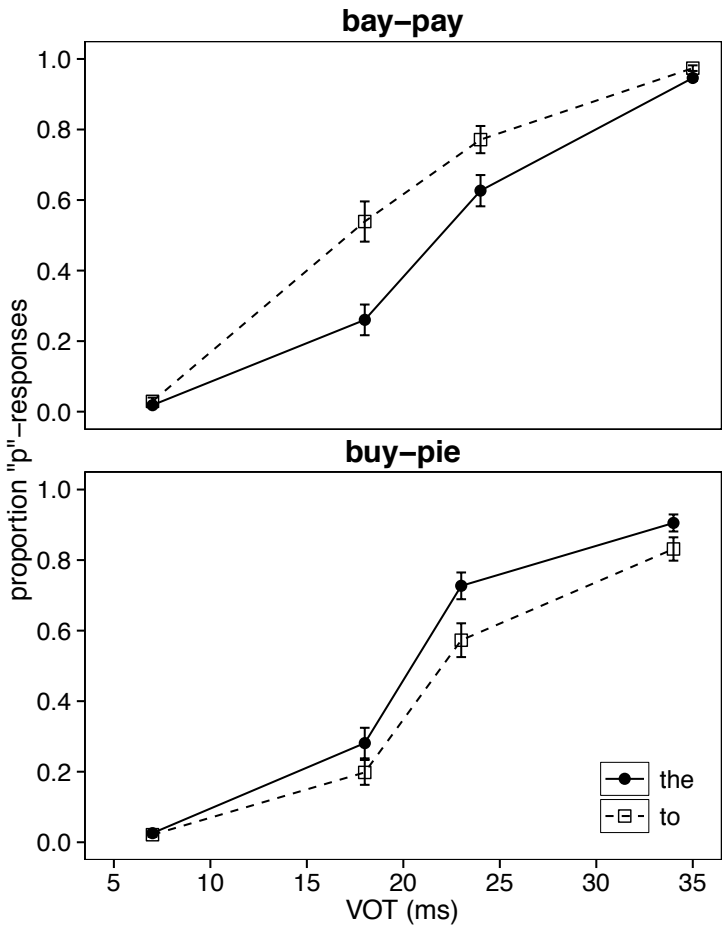

Figure 1.1. Mean proportion of /*p*/-responses to tokens from each VOT continuum in Experiment 1.1 after noun-biasing and verb-biasing sentence contexts. Error bars represent standard error.

The results of Experiment 1.1 are shown in Figure 1.1. Because this study was designed to examine contextual effects on the processing of ambiguous speech, data from responses to the two middle tokens in each continuum were analyzed. The mean proportion of /*p*/-responses for those intermediate tokens in each context and continuum

are summarized in Figure 1.2. Individual subjects' data were excluded if they did not make at least 10% /*b*/-responses and 10% /*p*/-responses to the ambiguous tokens in a given continuum. Based on this criterion, all 49 subjects perceived at least one of the continua ambiguously (36 for *bay–pay* continuum; 46 for the *buy–pie* continuum; 33 for both continua). Finally, 205 trials (1.31% of responses) with extreme reaction time (RT) values were removed prior to analysis (>3 standard deviations from the mean RT for a given subject/target/context).

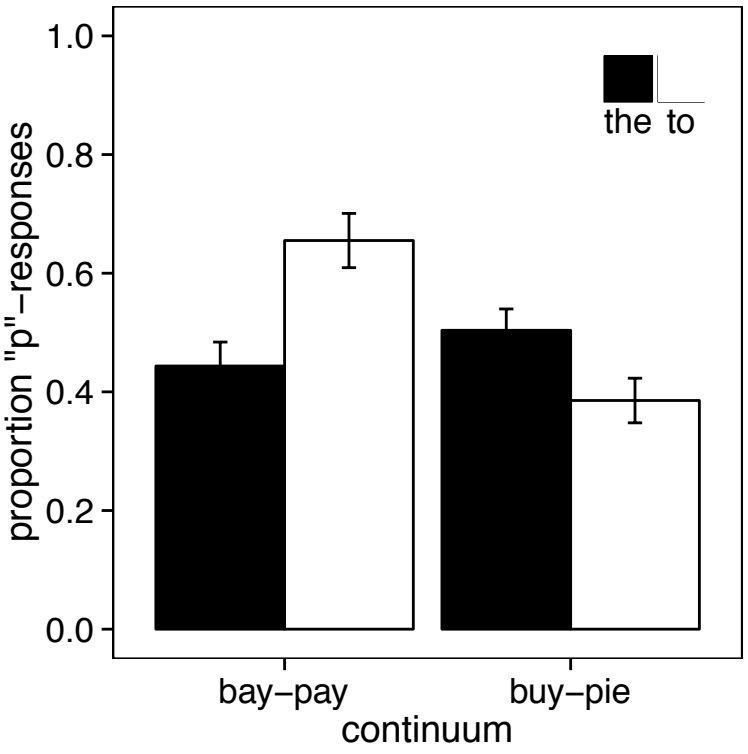

Figure 1.2. Mean proportion of /*p*/-responses to ambiguous tokens from each VOT continuum in Experiment 1.1 after noun-biasing and verb-biasing sentence contexts. Error bars represent standard error.

To test for an effect of sentential context on the identification of ambiguous stimuli, the data were analyzed using mixed effects logistic regression (Baayen, Davidson & Bates, 2008; Jaeger, 2008); a detailed description of the analyses can be found in the

Appendix C. The regression model included fixed effects for CONTEXT (verb-biased vs. noun-biased), CONTINUUM (*bay–pay* vs. *buy–pie*), and VOT (the VOT of each ambiguous token), along with all their two- and three-way interactions. All random intercepts and slopes were included for both subjects and items (i.e., main verbs; e.g., *hated*). Since the CONTINUUM factor crossed voicing (/*b*/ vs. /*p*/) and syntactic category bias (noun vs. verb), the critical test of a syntactic context effect is a CONTEXT × CONTINUUM interaction. A significant CONTEXT × CONTINUUM interaction would indicate that, after hearing *the* (which requires a noun rather than a verb), subjects were more likely to report hearing a */p/* if the target came from the *buy–pie* continuum, and more likely to perceive a */b/* if it came from the *bay–pay* continuum. As suggested by Figure 1.2's reversal in direction of the context effect within each continuum, the results showed a robust CONTEXT × CONTINUUM interaction ($\beta = 2.27$, SE = 0.30, $|z| = 7.53$, $p < 0.001$). Follow-up tests confirmed a crossover interaction between CONTEXT and CONTINUUM, indicated by a significant simple effect of CONTEXT on responses to stimuli from each continuum, but in opposite directions (*bay–pay*: $\beta = -1.37$, SE = 0.19, $|z| = 7.34$, $p < 0.001$; *buy–pie*: $\beta = 0.95$, SE = 0.20, $|z| = 4.82$, $p < 0.001$). All other effects that reached significance in the omnibus and follow-up analyses are reported and discussed in the Appendix C.

The central aim of Experiment 1.1 was to examine whether the obtained syntactic context effect was modulated by response latency. Thus, following the analysis procedure of Tuinman and colleagues (2014), we divided responses into two RT ranges (*fast* vs. *slow*) to test for differences in the size of this contextual bias. To do this, the responses of each participant within each cell of the experiment's design (20 responses for each subject/context/continuum/token, less outliers) were ranked according to their RTs. Then,

from each ranked list of RTs, the eight trials (40%) with the shortest RTs were labeled *fast* (mean RT = 577 ms, SD = 149 ms) and the eight trials (40%) with the longest RTs were labeled *slow* (mean RT = 1,012 ms, SD = 410 ms), omitting the mid-range. Together, the fast and slow RT ranges constituted a binary factor: SPEED.

Tuinman and colleagues (2014) showed that sentential context interacted with SPEED such that subjects' responses were less influenced by context at slow RTs than fast RTs. However, in the present study, CONTEXT (i.e., *the* vs. *to*) has the opposite effect on responses to stimuli from the *bay–pay* continuum than for stimuli from the *buy–pie* continuum. Therefore, the CONTEXT and CONTINUUM factors were recoded into a single factor (BIAS) with two levels (/*p*/-*congruent* vs. /*b*/-*congruent*), each corresponding to two types of trials. Trials were classified as /*p*/-*congruent* when a verb-biasing context (e.g., *Brett hated to...*) preceded a target from the *bay–pay* continuum (because a *pay*-response is congruent with the context in these trials) and when a noun-biasing context (e.g., *Valerie hated the...*) preceded a target from the *buy–pie* continuum (because a *pie*-response is congruent with the context in these trials). Conversely, trials were /*b*/-*congruent* if they contained a verb-biasing context and a target from the *buy–pie* continuum ("*...to* /*?ai*/") or a noun-biasing context and a target came from the *bay–pay* continuum ("*the* /*?ei*/"). Visually, the /*p*/-congruent trial-types correspond to the two conditions in Figure 1.2 with higher rates of /*p*/-responses, while the lower bars represent /*b*/-congruent trials.

Having created the SPEED and BIAS factors, we examined whether there was a BIAS × SPEED interaction similar to what was observed in previous work (Tuinman et al, 2014; *cf.* van Alphen & McQueen, 2001). Figure 1.3 shows the mean proportion of /*p*/-

responses subjects made for ambiguous tokens in /*p*/-congruent and /*b*/-congruent conditions in each of the RT ranges. A logistic regression analysis with fixed effects for BIAS, SPEED, VOT, their two- and three-way interactions, and all corresponding random intercepts and slopes for subjects and items was conducted (see the Appendix C for details). This analysis revealed a significant BIAS × SPEED interaction ($\beta = 0.51$, SE = 0.15, $|z| = 3.30$, $p < 0.001$), wherein responses were more likely to be syntactically congruent – and hence show a larger bias effect – in faster responses (61.6% /*p*/-responses to ambiguous tokens in /*p*/-congruent conditions vs. 41.4% in /*b*/-congruent conditions) than in slower responses (/*p*/-congruent: 55.6%; /*b*/-congruent: 42.0%). Additional effects that reached significance are reported and discussed in the Appendix C.

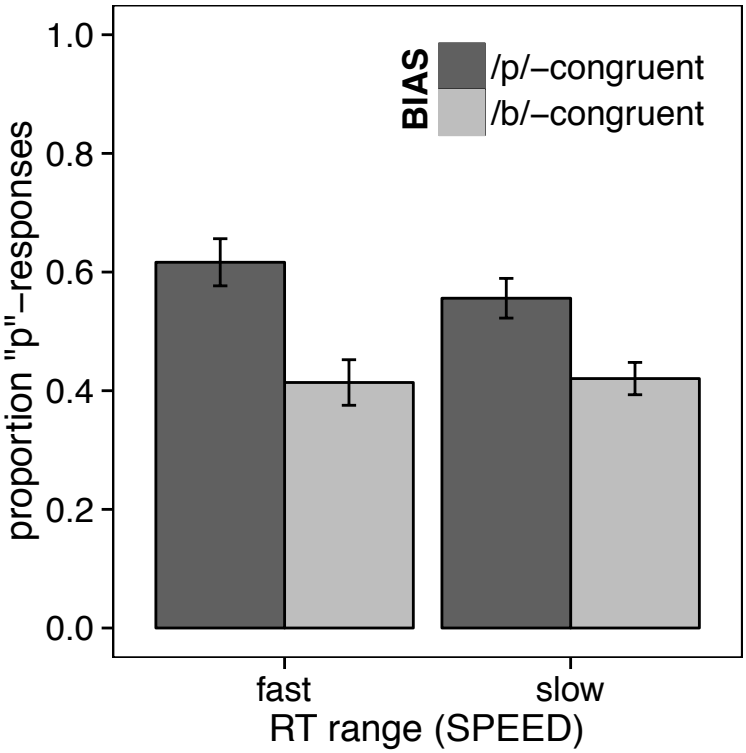

Figure 1.3. Mean proportion of /*p*/-responses in fast and slow responses to ambiguous tokens in /*p*/-congruent ("*...to* /***?****ei*/" and "*the* /***?****ai*/") and /*b*/-congruent ("*...the* /***?****ei*/" and "*to* /***?****ai*/") conditions in Experiment 1.1. Results indicate a weaker effect of BIAS in *slow* responses than in *fast* responses (see main text). Error bars represent standard error.

1.2.3. Discussion

Experiment 1.1 was designed to investigate one possible alternative explanation for previous results showing a diminishing influence of context on spoken word recognition (van Alphen & McQueen, 2001; Tuinman et al., 2014). Here, as in previous work, acoustic targets were manipulated to be phonetically ambiguous between two words and embedded in sentential contexts that rendered one of those words ungrammatical (or at least less plausible). However, unlike previous studies in which subjects could identify which response (button-press) would be congruent with the context before the target was presented, Experiment 1.1's design made it impossible for responses to the target stimuli to be systematically biased by the context if such responses

were planned prior to target presentation. Despite the addition of this control, the results of Experiment 1.1 showed that contextual biases on spoken word recognition still diminished over time, replicating earlier results.

Having ruled out this alternative explanation, the central theoretical question that remains is whether the observation of a weakening bias effect in Experiment 1.1 (and elsewhere; van Alphen & McQueen; Tuinman et al, 2014) is inconsistent with interactive speech perception models. As described earlier, this interpretation follows from the argument that top-down feedback permanently overwrites pre-lexical information in such models (e.g., TRACE). However, this argument rests on the critical assumption that subjects' decisions (both in previous experiments and in Experiment 1.1) tap into pre-lexical processing levels. In previous studies (van Alphen & McQueen, 2001; Tuinman et al., 2014), subjects made word identification decisions, which reflect the relative activation of competing *lexical* representations. Results reflecting lexical-level decisions cannot be taken as evidence either for or against top-down feedback to pre-lexical levels. Similarly, although Experiment 1.1 employed a phoneme identification task, it remains possible that subjects were monitoring for the four possible target words (*bay*, *pay*, *buy*, *pie*) and learning that each response button corresponded to two words, and thus were implicitly engaging in word identification. Experiment 1.2 aimed to resolve this potential issue by making lexical-level decisions difficult, if not impossible.

1.3. Experiment 1.2

Experiment 1.2 considered a second factor important for the interpretation of contextual influences that diminish in slower responses. Previous studies showing this pattern of results employed word identification tasks, not phoneme identification tasks

(van Alphen & McQueen, 2001; Tuinman et al., 2014). It is essential to consider the task-specific linking hypothesis (*cf*. Magnuson, Mirman & Harris, 2012; Tanenhaus, Magnuson, Dahan & Chambers, 2000) that transforms model activations into behavioral predictions in TRACE: "word identification responses are assumed to be based on readout from the word level, and phoneme identification responses are assumed to be based on readout from the phoneme level" (McClelland & Elman, 1986; p. 21). Thus, since subjects were identifying words (not phonemes), a model like TRACE would predict that lexical responses should reflect word-level (not phoneme-level) activations. Consequently, the time course of context effects demonstrated in previous work may not, in fact, be inconsistent with either interactive models (generally) or TRACE (specifically) because word identification data are not relevant to the question of the presence or absence of feedback from lexical to pre-lexical nodes.

Experiment 1.2 modified the design of Experiment 1.1 to discourage subjects from adopting word identification strategies. In particular, subjects performed a phoneme identification task in which the critical target stimuli from the *bay–pay* and *buy–pie* continua were embedded among twenty filler target words beginning with /b/ or /p/. We hypothesized that, when responding to 24 unique target words, subjects would have to monitor for the identity of a target's initial consonant in order to perform the phoneme identification task rather than utilizing word-level strategies. As such, subjects' responses in Experiment 1.2 should reflect the relative evidence for competing pre-lexical representations, a prerequisite to using any time course analysis to discriminate between models with and without interactive feedback.

The stimuli for Experiment 1.2 included sentences of the same form as Experiment 1.1 (e.g., *Brett hated to...*), but 160 sentences ending with critical targets (an acoustically manipulated token from either the *bay–pay* or *buy–pie* continuum) were embedded among 800 sentences ending with filler targets.

1.3.1. Methods

1.3.1.1. Materials

1.3.1.1.1. Critical Targets

The same 8 critical target tokens used in Experiment 1.1 were used in Experiment 1.2.

1.3.1.1.2. Filler Targets

Twenty words (10 beginning with /*b*/; 10 beginning with /*p*/) were selected to serve as filler targets in Experiment 1.2 (for a complete list, see Appendix B). The list of filler words included nouns (e.g., *bull*), verbs (e.g., *put*), and syntactically ambiguous words (e.g., *plan*). The syntactic bias of each filler word (rate of usage of the word as a noun vs. a verb) was computed using the Penn Treebank (Marcus, Marcinkiewicz & Santorini, 1993), and the list of filler words beginning with /*b*/ and /*p*/ were balanced for their average syntactic bias. Finally, of the twenty filler words, eight of them composed four minimal pairs (e.g., *bull*, *pull*) so that, when combined with the critical targets, half of the targets in Experiment 1.2 comprised minimal pairs. The same female speaker that recorded the stimuli for Experiment 1.1 read aloud sentences ending with the filler target words. The filler targets were then spliced out of the recorded sentences, scaled to have the same maximum volume as the critical targets, and appended to 20 ms of silence (like

the critical targets), but were not acoustically manipulated (e.g., by altering the VOT of onsets).

1.3.1.1.3. Sentence Contexts

Each critical and filler target was appended to each of the forty sentence contexts used in Experiment 1.1; for each of the twenty main verbs (that is, for each item; e.g., *hated*), there was a set of noun-biasing and verb-biasing contexts (*Valerie hated the...*, *Brett hated to...*) followed by every critical and filler target. Of the twenty item-sets, ten were randomly selected for each participant in Experiment 1.2, and that participant heard all sentence stimuli associated with those ten items (noun-biased and verb-biased sentences, ending in all critical and filler targets). Thus, the design remained fully within-subjects and within-items, although not every subject heard the same items. As in Experiment 1.1, critical stimuli included four tokens from each of the VOT continua. In an effort to balance the likelihood that subjects would hear a sentence that ended with any given word, participants heard each of their filler sentences twice over the course of Experiment 1.2. In all, subjects heard 800 filler trials (10 items * 2 levels of CONTEXT * 20 filler target words * 2 presentations) and 160 critical trials (10 items * 2 levels of CONTEXT * 2 levels of CONTINUUM * 4 tokens from each VOT continuum), yielding a total of 960 trials.

1.3.1.2. Participants

Twenty Brown University undergraduates who were self-reported native monolingual American English speakers with normal hearing received course credit to participate in Experiment 1.2. None had participated in Experiment 1.1 or any previous norming study.

1.3.1.3. Task

The task was identical to Experiment 1.1: all critical and filler sentences were presented binaurally in a random order after eight practice trials and participants were asked to indicate with a button-press whether the last word in each sentence began with a "b" or a "p" (response mapping was counterbalanced between subjects). All instructions were the same as in Experiment 1.1. Subjects were offered three breaks (after every 240 stimuli).

1.3.2. Results

Phoneme identification responses to trials ending with filler words were highly accurate (98.2% correct). This was unsurprising because all filler words were naturally produced tokens, so they were not phonetically ambiguous. Responses to critical trials from each continuum are shown in Figure 1.4. All further analyses followed the same approach as Experiment 1.1's analyses, independently fitting identical logistic regression models using subjects' responses to ambiguous tokens from the two VOT continua (additional details in the Appendix C). Using the same criterion as in Experiment 1.1, a subject's data were excluded if the intermediate tokens of a continuum were not perceived ambiguously (the responses of 18/20 subjects were included for at least one continuum: 15 for *bay–pay*; 13 for *buy–pie*; 10 for both). No trials warranted removal following an RT outlier analysis (same criteria as Experiment 1.1).

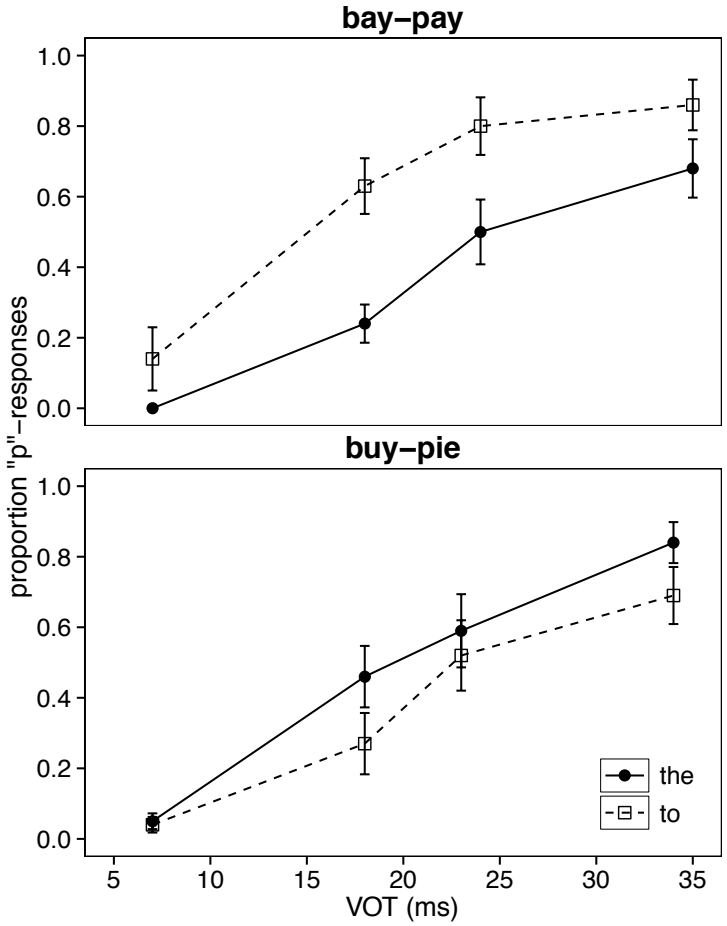

Figure 1.4. Mean proportion of /*p*/-responses to tokens from each VOT continuum in Experiment 1.2 after noun-biasing and verb-biasing sentence contexts. Error bars represent standard error.

The results of the first three-factor (CONTEXT × CONTINUUM × VOT) mixed effects regression revealed a significant CONTEXT × CONTINUUM interaction (β = 2.75, SE = 0.51, $|z|$ = 5.34, $p < 0.001$; see Figure 1.5), confirming that subjects' phonemic decisions about ambiguous targets were influenced by the grammaticality of targets given a preceding context. Follow-up tests confirmed the crossover interaction (opposite effects of CONTEXT in each continuum; *bay–pay*: β = -1.87, SE = 0.25, $|z|$ = 7.46, $p < 0.001$; *buy–pie*: β = 0.66, SE = 0.27, $|z|$ = 2.45, $p < 0.02$). Other effects are reported and discussed in the Appendix C.

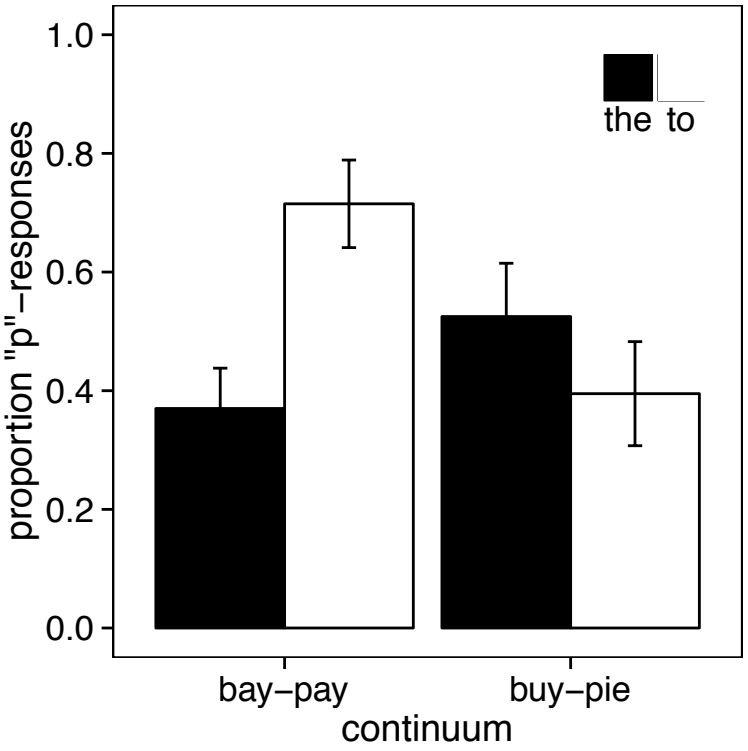

Figure 1.5. Mean proportion of /*p*/-responses to ambiguous tokens from each VOT continuum in Experiment 1.2 after noun-biasing and verb-biasing sentence contexts. Error bars represent standard error.

To test whether the influence of context on phoneme identification diminished over time, trials were recoded and split into two RT ranges (*fast*: mean RT = 616 ms, SD = 157 ms; *slow*: mean RT = 1,063 ms, SD = 423 ms) for a three-way BIAS × SPEED × VOT logistic regression, as in Experiment 1.1. This analysis provided no evidence of a BIAS × SPEED interaction ($p > 0.86$; see Figure 1.6). That is, the effect of the syntactic manipulation on the rate of /*p*/-responses did not diminish between faster responses (/*p*/-congruent: 60.6%; /*b*/-congruent: 40.6%) and slower responses (/*p*/-congruent: 61.3%; /*b*/-congruent: 41.9%). Other effects are discussed in the Appendix C.

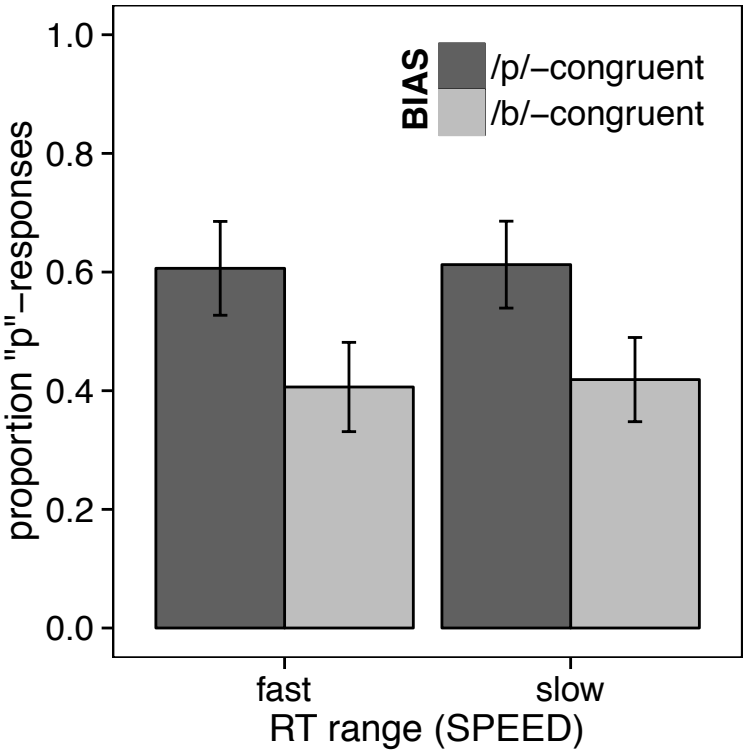

Figure 1.6. Mean proportion of /*p*/-responses in fast and slow responses to ambiguous tokens in /*p*/-congruent ("...*to* /***?***ei*/" and "*the* /***?***ai*/") and /*b*/-congruent ("...*the* /***?***ei*/" and "*to* /***?***ai*/") conditions in Experiment 1.2. Unlike Experiment 1.1, the BIAS effect in Experiment 1.2 is as strong in *slow* responses as in *fast* responses (see main text). Error bars represent standard error.

1.3.3. Discussion

The goal of Experiment 1.2 was to examine contextual influences on the identification of ambiguous targets using a task designed to elicit phonemic decisions. Of particular interest was the time course of such context effects, and especially whether these results would differ from previous word identification experiments (van Alphen & McQueen, 2001; Tuinman et al, 2014) and from Experiment 1.1, in which it was unclear whether subjects were engaging in word or phoneme monitoring strategies. Experiment 1.2, like Experiment 1.1, showed that syntactic context has a robust effect on subjects' responses (including their fastest responses). However, unlike Experiment 1.1, Experiment 1.2 showed that this contextual bias on phoneme identification was as strong

in slow responses as it was in fast responses. Van Alphen and McQueen (2001) argue that if the fast-arising bias in phoneme responses was the result of lexical feedback to pre-lexical representations (as hypothesized by interactive models), then "there should have been a similar shift (if not a stronger one as more time elapsed with more feedback) in slow responses" (p. 1069). Indeed, results of Experiment 1.2 are consistent with these predictions and thus support the view that pre-lexical representations are modulated by lexical feedback, as hypothesized by interactive models. Additionally, the results of Experiment 1.2 suggest that tasks in which word identification decisions are required or may be used strategically by participants may not tap pre-lexical representations.

1.4. General Discussion

The present work aimed to evaluate the validity of claims that the time course of context effects on speech perception is incompatible with interactive models of speech perception (*cf.* van Alphen & McQueen, 2001; Tuinman et al., 2014). According to this view, context effects within an interactive system should remain stable or grow over time, but they should not become weaker, because biasing feedback from lexical representations irreversibly overwrites an initially ambiguous representation of the acoustic input. Once pre-lexical representations are altered by top-down modulation, there is no recovering the record of the ambiguous signal, so the size of the bias effect should not diminish. Crucially, this logic is predicated on the assumption that subjects' responses reflect activation levels of pre-lexical representations, an assumption that Experiments 1.1 and 1.2 examined more closely.

Two key results emerged from Experiments 1.1 and 1.2. Firstly, as already discussed, the results of Experiment 1.2 suggest that when an experimental task is

designed to tap into pre-lexical processing, contextual influences on speech perception are robust and persistent over time. Secondly, the biasing effects from a preceding sentential context arose very rapidly in both experiments. In the present studies, the button-press that would represent a contextually congruent response depended on the target stimulus itself, so it was impossible for a crossover interaction to emerge unless subjects waited to hear the target stimuli. In other words, the fact that subjects' fast responses were biased suggests that the processing of ambiguous speech is influenced by the grammatical properties of competing words, that this influence is virtually immediate, that this rapid influence is not attributable to pre-planned responses, and that top-down expectations rapidly propagate to bias both lexical and pre-lexical representations. As such, our data provide strong evidence for immediate top-down effects from sentence context effects on speech perception. Taken together, these results suggest that the time course of contextual effects on phoneme recognition does not, in fact, challenge the interactive modeling framework. Next, we consider the extent to which these results actually support interactive models, and how they constrain autonomous models.

1.4.1 Implications for Interactive Models of Speech Perception

Despite the fact that subjects responded to the same stimuli in Experiments 1.1 and 2, the time course of the contextual bias effect differed between experiments, with Experiment 1.1's results matching the pattern obtained by word identification tasks (van Alphen & McQueen, 2001; Tuinman et al., 2014). An important question is why lexically-driven and phonemically-driven responses would generate different patterns with respect to the time course of context effects. Recall that in TRACE, outputs in *n*-alternative forced-choice tasks (such as phoneme monitoring/identification or visual

world eye-tracking) are generated probabilistically from among a set of alternatives that is identified depending on the task and stimuli (*cf*. Luce, 1959). TRACE's decision model keeps track of a running average of activation levels of the alternatives, which are nodes in a single layer of the model (*cf*. McClelland & Rumelhart, 1981). For instance, in a phoneme identification task, two units in the Phoneme layer (e.g., /*b*/ and /*p*/) constitute the output alternatives that are tracked (McClelland & Elman, 1986; McClelland, 1987), while in word recognition or visual world eye-tracking tasks, nodes in the Word layer (e.g., *bear* and *pear*) are identified as the output alternatives (e.g., Allopenna, Magnuson & Tanenhaus, 1998; Dahan, Magnuson & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus & Hogan, 2001; Magnuson, Dixon, Tanenhaus & Aslin, 2007; Magnuson, Tanenhaus, Aslin & Dahan, 2003; McMurray, Tanenhaus & Aslin, 2002). Since different tasks dictate that outputs reflect activation dynamics in different layers of the model, it is not surprising to observe unique patterns of results for word vs. phoneme identification tasks.

Nevertheless, TRACE (or any other model) still must explain why Word-level activations become less biased over time even though biased Phoneme-level activations persist. This question can only be fully addressed once models of speech perception incorporate sentence-level representations (*cf*. Strand, Simenstad, Cooperman & Rowe, 2014). However, it seems likely that an interactive model could capture this difference. A model with a relatively strong stabilizing force (i.e., decay) at the lexical level or a quickly decaying influence of lexical expectation from "supra-lexical" levels of representation would predict a transient context effect in word identification responses. Notably, Word-level decay is the strongest decay parameter in TRACE (McClelland &

Elman, 1986). Meanwhile, as van Alphen and McQueen (2001) suggest, pre-lexical representations that are modulated by lexical feedback may not recover from the top-down biasing influence (or at least may not recover as quickly as the lexical representations), leading to more persistent context effects on phoneme identification responses (as in Experiment 1.2).

1.4.2. Implications for Autonomous Models of Speech Perception

It is less clear to what extent the results of Experiment 1.2 challenge autonomous models of speech perception. Van Alphen and McQueen (2001) note that "if...sentential context effects are the result of a decision bias, predictions about their time course are much less clear" (p. 1059). As they acknowledge, their verbal account (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Magnuson, Mirman & Harris, 2012; Mirman & Britt, 2013) of sentential context effects on word recognition could predict many patterns of results, including the one their experiments suggest. Nonetheless, the present results do challenge one theoretical proposal they offer. In their data, responses in the slowest RT range failed to show a significant effect of preceding context. They ultimately interpret this as consistent with a model in which context effects on subjects' decisions are time-limited such that sentence-level processing can only bias responses while the syntactic parse of the sentence remains ambiguous (see also Mattys, Melhorn & White, 2007). However, the persistent top-down effects on subjects' responses in Experiment 1.2 challenges the view that sentence context has a time-limited effect (see also Bicknell, Jaeger & Tanenhaus, 2015; Bicknell, Tanenhaus & Jaeger, 2015; Connine, Blasko & Hall, 1991; Szostak & Pitt, 2013). In fact, even when the diminishing bias effect was replicated (in Experiment 1.1 and by Tuinman et al, 2014), the slow responses were still

significantly biased. Thus, any autonomous model of speech perception must be able to account for long-lasting contextual biases on both phoneme and word identification responses.

1.4.3. Predicting Behavior with Computational Models of Speech Perception

The present work underscores the need for explicit, testable computational models bridging speech perception and sentence processing, while also highlighting the importance of appropriately interpreting existing models' predictions. Neither TRACE nor Merge makes any predictions about sentential context effects. Indeed, McClelland and Elman (1986) conceded this point when they introduced TRACE, explicitly leaving the question to future research: "We have not yet included...higher level contextual influences in TRACE, though of course we believe they are important" (p. 60). In order to understand the mechanisms underlying speech perception in naturalistic environments, it is necessary to develop and test models in which spoken sounds/words are accompanied by a rich linguistic and non-linguistic context.

1.5. Conclusion

Interactive and autonomous models represent two powerful theories, each of which are capable of explaining many data regarding speech perception. Despite decades of arguments, rejoinders, clarifications, and revisions, they have proven difficult to distinguish. Given the persistence of the debate and the considerable power of each modeling framework, some have wondered whether any unique predictions exist that might settle the question (Cutler et al., 1987; Pitt & Samuel, 1993). The goal of this study was to assess whether the time course of context effects on speech perception is incompatible with interactive models, as has been proposed. Our results suggest that this

claim is unwarranted. Indeed, when an experiment is designed to tap into pre-lexical processing dynamics, the persistent influence of top-down lexical feedback can be observed. Our findings, therefore, indicate that there is even less evidence against interactive models of spoken word recognition than previously thought. On the other hand, while the present results provide some constraints for autonomous models, they cannot entirely rule out such a framework. Ultimately, it is clear that the lack of overt, testable, divergent predictions represents a fundamental barrier to resolving this long-running theoretical debate. Given this critical gap, we believe that the development of well-constrained, explicit computational models must play a central role in future research investigating the mechanisms underlying spoken word recognition.

1.6. Overview of Next Steps

Unfortunately, as noted earlier, neither TRACE nor Merge is designed to explicitly model the role of sentential context in spoken word recognition. In fact, as we will discuss in Chapter 2, despite strong evidence illustrating context effects (as seen in the experiments reported here in Chapter 1), the influence of sentential context on speech perception is poorly characterized in existing psycholinguistic models. In order to address this major gap, we now turn to the task of developing a computational model of speech perception capable of explaining top-down effects from a word's sentence context. As we will further show in Chapter 3, the model developed in Chapter 2 also provides a straightforward, natural account for the enormous variability in the size of top-down effects across different subjects, stimuli and experiments – an issue that has been almost completely ignored by previous models of speech perception. Finally, as we will discuss in Chapters 3 and 4, developing such a model promises to generate novel, testable

predictions that can elucidate properties of the cognitive and perceptual processing underlying auditory language comprehension in both healthy adults and patients with brain damage.

Chapter 2

Bayesian Integration of Acoustic and Sentential Evidence in Speech:

The BIASES Model of Spoken Word Recognition in Context

2.1. Introduction

2.1.1. Brief Introduction

In order to comprehend spoken language, a listener must ultimately map a perceived acoustic waveform onto some meaningful interpretation of the speaker's message. The processing that underlies this complex phenomenon can be thought of as consisting of at least three subroutines: pre-lexical speech processing, spoken word recognition, and auditory sentence comprehension. While all three of these are, of course, critical for arriving at an accurate, contextualized understanding of the meaning behind some speech that reaches a listener, spoken word recognition occupies a critical juncture between the lower-level perceptual processing of a signal and the higher-level processing by which listeners recruit linguistic and world knowledge to construe the meaning of the words they identify. Because of the crucial role words play as the gatekeepers of meaning, characterizing the computations involved in the recognition of spoken words is of critical importance (for recent reviews, see Mattys, 2013; Maguson, Mirman & Myers, 2013; Tanenhaus, 2007).

The fundamental computational problem (Marr, 1982) that must be solved in order to recognize words in speech is to infer which word (or sequence of words) was most likely to have been produced by the speaker. This decoding of the perceived speech stream would be trivial if words were consistently produced with a single acoustic form, any given acoustic signal corresponded to exactly one word, and perception always took

place under noise-free conditions. However, none of these facts is true of speech perception in the real world. Quite to the contrary, the number of ways in which noise, ambiguity, and uncertainty can compromise processing is daunting.

Nonetheless, despite all of the potential barriers to successful mapping of signal to meaning, healthy listeners rarely experience difficulties in speech processing. Even when speech is intentionally degraded in the laboratory to tax the perceptual system, sounds and words are typically perceived with high accuracy (e.g., Luce & Pisoni, 1998; Cutler, Weber, Smits & Cooper, 2004). Indeed, listeners often fail to notice when a segment in a word is replaced with white noise or a cough (Warren & Obusek, 1971), and, when they do, it seldom impedes comprehension.

What accounts for this robustness in the face of such pervasive degradation? A complete understanding of this issue requires, minimally, answering two fundamental questions: what cues are available to the listener, and how are these cues leveraged in order to overcome the ambiguity inherent in the input. Broadly, investigations of the first question – which cues do listeners utilize during speech recognition – have shown that listeners integrate both *bottom-up* (sensory-based) and *top-down* (knowledge-based) cues (for review, see Samuel, 2011). That is, in addition to leveraging bottom-up acoustic cues derived from perceptual processing of the speech signal such as voice-onset time and formant values, listeners also exploit cues that require higher-level cognitive processing of the signal. For instance, listeners are sensitive to whether or not different potential interpretations (e.g., *goat* vs. *coat*) of some speech input are sensible given the preceding sentential context (e.g., *The busy farmer hurried to milk the...*) (Borsky, Tuller & Shapiro, 1998).

The second question – how the various cues are integrated and come to influence speech recognition – is the focus of the present work. This wide-ranging question has motivated a host of theoretical and computational models focusing on different aspects of speech processing from how multiple distinct bottom-up cues are weighted (e.g., Massaro & Oden, 1980; Nearey, 1990, 1997; Oden & Massaro, 1978; Repp, 1982, 1983; Toscano & McMurray, 2010) to how multisensory information sources are combined (e.g., Diehl & Kluender, 1989; Fowler & Rosenblum, 1991; Kluender, 1994; Massaro, 1987; McGurk & MacDonald, 1976; Ostrand, Blumstein & Morgan, 2011; Rosenblum, 2005) to what mechanisms give rise to lexical biases on word recognition (Elman & McClelland, 1988; McClelland, Mirman & Holt, 2006; McQueen, Norris & Cutler, 2006; McQueen, Jesse & Norris, 2009; Norris, McQueen & Cutler, 2000). However, one major theoretical gap that remains concerns how top-down information available from a sentence context is integrated with bottom-up cues. This gap is especially conspicuous given that everyday speech rarely features words produced and perceived in isolation, and sentential context has consistently been shown to impact the recognition of spoken words (e.g., Borsky et al, 1998; Lieberman, 1963; Warren & Warren, 1970). The present work aims to narrow this gap by advancing one approach to modeling the integration of top-down and bottom-up cues during speech perception. In particular, we argue that viewing speech perception through the lens of Bayesian cue integration provides a powerful, principled framework to understand a wide range of behavioral data. To this end, we outline the issues addressed in this chapter, which is organized into three parts.

2.1.2. Overview of Chapter 2

First, we address the question of why this gap exists at all. We discuss several practical and theoretical bottlenecks associated with representing sentential context and modeling the mechanisms by which context might come to influence spoken word recognition.

Second, we argue that these challenges motivate the specification of a computational-level (Marr, 1982) model of spoken word recognition capable of explicitly integrating bottom-up and top-down information sources. We present a Rational Analysis (Anderson, 1990) of spoken word recognition in context and propose that a Bayesian modeling approach may offer key insights into the information processing that underlies spoken language processing.

Third, we introduce *BIASES* (short for *Bayesian Integration of Acoustic and Sentential Evidence in Speech*), a Bayesian model of spoken word recognition in context. BIASES is a novel, flexible computational framework for simulating human behavior in word recognition tasks and for testing psycholinguistic theories about how bottom-up and top-down information sources are represented and integrated by listeners. Adopting a model like BIASES involves embracing three basic assumptions: (1) that listeners are sensitive to fine-grained acoustic properties of spoken words; (2) that they are also sensitive to fine-grained differences in the chances of encountering different words in a given sentence context; and (3) that, when identifying spoken words, they integrate these information sources with consideration for the relative reliability of each available cue. We review the robust empirical evidence that supports these assumptions, and, in turn, the Bayesian approach to spoken word recognition.

2.2. Sentential Context and Connectionist Models of Spoken Word Recognition

Existing computational models of spoken word recognition have not directly addressed how a word's processing is influenced by its sentential context. Before motivating the present model, it is worth reviewing the evidence under consideration and possible explanations for why current models fail to account for this evidence.

2.2.1. Modulation of Spoken Word Recognition by Sentential Context

Several decades of research have made it clear that the recognition of a spoken word is not independent of its context. Words that are unintelligible when presented in isolation can be readily identified in context (Lieberman, 1963; Pickett & Pollack, 1963; Hunnicutt, 1985; Fowler & Housum, 1987), and prior exposure to an acoustically clear prime sentence improves listeners' recognition of conceptually related words in a subsequent acoustically degraded sentence (Guediche, Reilly & Blumstein, 2014). Moreover, in addition to facilitating recognition of speech in noise, a word's context can shape a listener's interpretation of spoken words that are ambiguous between two or more words in their language. For instance, words that have been digitally manipulated to replace a critical speech segment with extraneous non-speech acoustic material (such as a cough or white noise) are more often recognized as words that are consistent with the context in which the word was presented: /**il*/, where * represents the digitally substituted non-speech sound, may be identified as *wheel*, *heel*, *peel* or *meal* depending on other words appearing in the same sentence (e.g., *axle*, *shoe*, *orange* or *table*) (Warren & Warren, 1970; Warren & Sherman, 1974).

Furthermore, related evidence shows that such contextual effects are not restricted to the restoration phonetic information that is missing altogether. This line of work utilizes fine-grained acoustic manipulations of phonetically relevant parameters of natural

speech tokens to render the resulting stimuli ambiguous between two possible words. When these stimuli are presented in sentences that are consistent with one word or the other, they tend to be perceived as the word that is congruent with the context in which it is embedded. For example, subjects are more likely to identify a phonetically ambiguous stimulus between *goat* and *coat* as *goat* when it follows a sentence like *The busy farmer hurried to milk the...* that when after sentences like *The careful tailor stopped to button the...* (Borsky, Shapiro & Tuller, 1998). Such biases have been widely corroborated, whether the manipulated contextual constraints operate at the semantic (Borsky et al, 1998; Garnes & Bond, 1976; Miller, Green & Schermer, 1984; Connine, 1987; Guediche, Salvata & Blumstein, 2013), syntactic (Fox & Blumstein, in press; Tuinman, Mitterrer, & Cutler, 2014; van Alphen & McQueen, 2001), morphological (Martin, Monahan & Samuel, 2011), or pragmatic level (Rohde & Ettlinger, 2012; Do, 2011).

With such strong evidence for contextual effects on spoken word recognition, it is somewhat surprising that word recognition models have thus far offered no explicit account for these data. The treatment of sentential context in most existing spoken word recognition models can generally be classified into four categories. First, some models ignore the role of sentential context, focusing on other aspects of spoken word recognition (e.g., PARSYN: Luce, Goldinger, Auer & Vitevitch, 2000; ARTWORD: Grossberg & Myers, 2000; Merge: Norris, McQueen & Cutler, 2000; LAFF: Stevens, 2002). A second group of models explicitly leave the question to future research (e.g., LAFS: Klatt, 1979; TRACE: McClelland & Elman, 1986; MINERVA 2: Goldinger, 1998; Hintzman, 1986). A third set of models asserts that incorporating sentential context would be a "straightforward" extension of the more basic model (e.g., NAM: Luce &

Pisoni, 1998; Shortlist: Norris, 1994; Shortlist B: Norris & McQueen, 2008; SpeM: Scharenborg, Norris, ten Bosch & McQueen, 2005; see also Norris, McQueen & Cutler, 2015). Finally, there are some theories about what role sentential context might play in speech recognition have been presented (e.g., Logogen: Morton, 1969; Race: Cutler & Norris, 1979; Cohort: Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978). Several members of this set – most notably the Cohort model – were prominent theories that guided early research on spoken word recognition, and, although they have been abandoned in light of empirical challenges to some of their specific claims, the principles they embodied (e.g., graded activation, competition, autonomous vs. interactive model architectures) remain influential today.

However, with respect to the subject of the present work – sentential influences on spoken word recognition – this fourth group exemplifies what has probably been the most common treatment of the issue. That is, many theories have relied on "verbal models" which might explain some aspects of processing that is likely implicated (or, just as often, what sorts of processing might be precluded; *cf.* Shillcock & Bard, 1993; Tanenhaus, Leiman & Seidenberg, 1979; Tanenhaus & Lucas, 1987) during sentence-level speech processing. However, these models have a number of disadvantages, chief among them being that they are often incompletely specified. While verbal models are critical to theory development and are useful for generating and testing many predictions, it is often difficult or impossible to assess a theory's adequacy or viability if it is not mathematically or computationally implemented, and it is even more difficult to compare its predictions to another competing theory (see, e.g., Magnuson, Mirman & Harris,

2012). In short, theories and models falling into the third and fourth categories above leave much work to be done.

The lack of a comprehensive model of spoken word recognition in context is probably attributable to a number of factors. For one, many difficult and important questions can be (and have been) explored without the additional complication of modeling what are logically more abstract representations and cognitive functions (e.g., the composition of meaning). However, the exclusion of sentence-level information in existing models of speech perception can also be traced to major challenges presented by the predominant modeling approach to examining higher-level influences on word recognition.

2.2.2. Challenges in Modeling Context Effects on Spoken Word Recognition

Many of the most influential models of spoken word recognition, including TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and Merge (Norris, McQueen & Cutler, 2000), are based on interactive activation networks (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1981, 1982). In such models, cognitive representations are connected to one another in a network, with each representation characterized by some amount of activation. Activation propagates through the network as a function of the connections between representations and the sensory input presented to the network. In localist connectionist models of spoken word recognition, each representation is designated by a node that stands for a linguistically-relevant unit (e.g., a word or a phoneme), and these nodes are organized into layers (*cf.* McClelland & Rumelhart, 1986; Page, 2000). The nodes within a given layer represent mutually exclusive hypotheses (*cf.* Smolensky, 1986) about which linguistic units might be

(underlyingly) present within a given speech signal. For instance, the words *goat* and *coat* would be represented as unique nodes in the Word layer of a model because a spoken word may be an exemplar of *goat* or *coat*, but not both. Although the exact details of how units are connected differ from model to model, nodes within a layer are typically connected via inhibitory connections, while mutually consistent linguistic units (e.g., a word-initial /*g*/ in the Phoneme layer and *goat* in the Word layer) would be connected via excitatory connections. Critically, though, if some node A is connected to some other node B via an excitatory connection, then when node A increases in activation, node B will also tend to increase in activation. On the other hand, if the connection from node A to node B is inhibitory, then when node A increases in activation, node B will tend to decrease in activation (for a recent review of interactive activation models in speech perception, see McClelland, Mirman, Bolger & Khaitan, 2014).

2.2.2.1. Challenges of Modeling Context Effects: Representing Context

Given this modeling framework, it is not immediately clear how one should incorporate sentential context. Perhaps the most obvious question is: how should sentence-level information be represented? In order to capture semantic context effects (e.g., more *GOAT* responses after sentences about *milking* than *buttoning*), semantic relationships among words must somehow be incorporated into the model. This might be made possible by constructing a layer of semantic features that has excitatory connections to some nodes in the Word layer based on the words' meanings (Chen & Mirman, 2012; Cree, McRae, McNorgan, 1999; Rogers & McClelland, 2004). Alternatively, it might be possible to ignore semantic features altogether and, instead, connect word units such that words can excite other related words based on semantic associativity norms (Deerwester,

Dumais, Fornas, Landauer & Harshman, 1990; Dumais, 2004; Landauer & Dumais, 1997) or other measures (Fellbaum, 1998; Miller, 1995; Miller, Beckwith, Fellbaum, Gross & Miller, 1990; Miller & Fellbaum, 1991). However, it is not clear which of these possibilities (or what other solution) is more consistent with the organization of listeners' semantic knowledge (*cf.* Andrews, Vigliocco & Vinson, 2009; De Deyne & Storms, 2008; Riordan & Jones, 2011; Steyvers & Tenenbaum, 2005), nor is it clear from existing data which model would best explain context effects in the domain of word recognition. Moreover, while some such model enhancement might be able to capture semantic context effects, explaining syntactic and pragmatic context effects would require the addition of more connections and/or layers (*cf.* McClelland, St. John & Tarban, 1989; Recchia, Sahlgren, Kanerva & Jones, 2015; Rohde, 2002; St. John & McClelland, 1990; Strand, Simenstad, Cooperman & Rowe, 2014).

2.2.2.2. Challenges of Modeling Context Effects: Activation Dynamics

Even if this the issue of representation could be solved, it is not straightforward to merely add more connections to the architecture of an existing activation-based model because it is not clear how contextual information should come to influence words' activation levels. Adopting the same activation dynamics assumptions used in existing models, the effect of sentential context might be excitatory (for words that are supported by the context). On the other hand, the effects could be implemented via inhibitory connections, essentially ruling out words that are not supported by the context (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978; for a similar approach in spoken word production, see Dell, Oppenheim, & Kittredge, 2008). Alternatively, rather than directly altering words' activation levels, sentential context might induce adjustments to

words' *propagation threshold* (a parameter governing the amount of activation required before a node's activation begins to influence other nodes in the network) or their *activation gain* (a parameter governing how easily words become more activated).

There is no *a priori* reason to believe that one of these mechanisms is more likely than any other, so additional assumptions and free parameters are needed. Furthermore, what works for modeling semantic context effects, where contextual cues (e.g., *milked*) tend to support specific words (e.g., *goat*), may not be able to capture syntactic context effects, where contextual cues (e.g., *the*) tend to support categories of words (e.g., *nouns*), but not specific words (see Fox & Blumstein, in press).

2.2.2.3. Challenges of Modeling Context Effects: Representing Time

A third issue that makes it difficult to model sentential influences on word recognition in connectionist models arises from the way activation dynamics transpire over time as the speech signal unfolds. Recall that nodes in the same layer are generally considered to be mutually inhibitory: when more than one word is partially activated, the most activated representation(s) tends to crowd out other active nodes (Thomas & McClelland, 2008). This architectural feature is almost universally true of the Word layers of localist spoken word recognition models (Gaskell, 2007), as it allows activation-based models to account for competition among multiple lexical candidates during recognition (Allopenna, Magnuson & Tanenhaus, 1998; Andruski, Burton & Blumstein, 1994; Frauenfelder & Floccia, 1998; Gaskell & Marslen-Wilson, 1999, 2002; Magnuson, Dixon, Tanenhaus & Aslin, 2007; McMurray, Tanenhaus & Aslin, 2002; McMurray, Tanenhaus, Aslin & Spivey, 2003; McMurray, Tanenhaus, Aslin, Spivey & Subik, 2008; McQueen, Norris & Cutler, 1994; Norris, McQueen & Cutler, 1995; Righi, Blumstein,

Mertus & Worden, 2009; Utman, Blumstein & Burton, 2000; Vitevitch & Luce, 1998, 1999).

However, this crucial architectural feature becomes problematic when the model is scaled up in order to account for sentential context effects, where multiple words *should* become activated in sequence. In the Merge model (Norris et al, 2000), for instance, activation of a lexical representation at one point in time will tend to suppress activation levels of other words at future points in time. Even if segmentation of the speech signal into words is taken for granted, it is clear that adapting Merge to account for sentential context effects would depend on how context is represented and how it comes to influence words' activation levels.

TRACE (McClelland & Elman, 1986), on the other hand, deals with time in a very different way, replicating the entire network for each time slice, so the inhibitory word-word connections only act on words that begin at the same point in time. From its inception, the implausibility of this aspect of TRACE's architecture has been widely acknowledged by the model's proponents and opponents alike (McClelland, Mirman & Holt, 2006; Norris, 1994) because of the enormous number of nodes required to model continuous speech. Therefore, modeling context effects by adding additional connections between associated words at different time-points or adding additional layers replicated for every time-point along with the rest of TRACE, would only exacerbate this problem.

Meanwhile, Shortlist's (Norris, 1994) representation of time is based on time-delayed recurrent neural networks (Elman 1990; Norris, 1988, 1990, 1993), which occupy a middle ground between the drawbacks of Merge and TRACE. Still, Norris' (1994) brief discussion of how Shortlist might be adapted in order to account for the

various sentential context effects observed in the literature does not directly address either the representation of context or how the dynamic modulation of activation in the time-delayed network would function.

2.2.2.4. Context Effects Without Connectionist Models

While all of these issues stem from important questions about spoken word recognition, they also pose significant challenges that even the most successful existing models have, so far, not addressed. Ultimately, however, these hurdles arise directly from the first choice made at the inception of each model: the choice to adopt a connectionist framework. Existing models were designed to address the architecture of a system supporting isolated word recognition, and adapting these architectures to solve a distinct computational problem – recognizing spoken words within the rich context of natural language – is a limiting approach when it comes to modeling context effects. As an alternative, it is possible to characterize the information processing architecture that must underlie any explanation of sentence-level influences on the recognition of spoken words, while acknowledging that there might be many possible architectures (representational systems, activation/dynamical assumptions, and implementations of time-varying input and processing) that could achieve the necessary computations (Marr, 1982). In the present work, we take this alternate path, analyzing the computational problem associated with word recognition in context from the beginning.

2.3. A Computational-Level Analysis of Spoken Word Recognition

A useful starting point for a computational analysis of spoken word recognition is with a *rational analysis* (Anderson, 1990), wherein a cognitive system is considered with respect to the system's goals, the environment in which the system must operate, and the

computational limitations of the system (Anderson, 1991). Anderson's (1990) Principle of Rationality presumes that a cognitive system is optimized with respect to these factors, so, to the extent that some data do not fit the rational model's predictions, these discrepancies will suggest that the modeler's original assumptions about the system's goals, environment, or limitations were inaccurate. These inconsistent data, in turn, guide the updating of a rational model's initial assumptions.

2.3.1. Bayesian Models of Spoken Word Recognition

Recent years have seen a notable rise in the application of rational analysis to questions in speech perception. Feldman, Griffiths and Morgan (2009) presented a rational analysis of speech sound perception and categorization, showing that the listeners' discrimination and perceptual classification of vowel tokens could be explained by assuming their behavior reflected optimal (Bayesian) perceptual inference under uncertainty. In the same vein, Shortlist B (Norris & McQueen, 2008) exemplifies the rational analysis approach to spoken word recognition, accounting for several classic effects in the psycholinguistic literature without appealing to the notion of activation at all. Although their details differ, and although only the latter model focuses on word recognition specifically, both models follow the same basic logic. Similarly, the present model of spoken word recognition in context also follows this logic, so we now turn to an outline of the foundational principles these models share.

Any rational analysis of a cognitive system must begin by identifying the goal of the system. Following Norris and McQueen (2008), and as suggested at the outset of this chapter, we take the purpose of the speech recognition system to be the recovery of the word (or words) produced by the speaker. For the purpose of exposition, we limit the

present discussion to a special case wherein the listener's goal is to infer the most likely single word given the perceived speech signal.

How would a rational system achieve this goal? If there is only one word that could possibly have produced the perceived signal, then the optimal decision is obvious: if all other words have been ruled out, then the signal must be an exemplar of the only remaining option (Doyle, 1890). However, since such certainty is often elusive when recognizing words in the real world, how should a rational system select the most probable word given incomplete information?

Under this view, spoken word recognition amounts to a specific case of a more general problem: inference under perceptual uncertainty. All such computational problems share the same mathematically optimal solution, which is defined by the ideal observer framework (Geisler, 2003; Geisler & Kersten, 2002). An ideal observer is one that always makes the best possible guess when identifying the likely source of some observed data, and its behavior is given by Bayes' rule (Knill, Kersten & Yuille, 1996). According to Bayes' rule (Equation 2.1), for any exhaustive set of mutually exclusive hypotheses H, the probability that any given hypothesis h_i in H is true, given some observed data d, is given by:

Equation 2.1

$$p(h_i|d) = \frac{p(d|h_i)p(h_i)}{\sum_{h_j \in H} p(d|h_j)p(h_j)}$$

Because the denominator of the right side of Bayes' rule is constant over all h_j in H, Bayes' rule is often stated in its proportional form (Equation 2.2):

Equation 2.2

$$p(h_i|d) \propto p(d|h_i)p(h_i)$$

The key principle embodied by Bayes' rule is that, having observed *d*, the so-called *posterior probability* of a given alternative, $p(h_i|d)$, depends on two general classes of information: how representative of that alternative *d* is, and on how probable the alternative (h_i) was in the first place. These two pieces of information are referred to, respectively, as an alternative's *likelihood*, $p(d|h_i)$, and its *prior probability*, $p(h_i)$. An ideal observer integrates these two sources of information by computing the posterior probability for each alternative in *H*, and ultimately selecting the alternative with the largest posterior probability.

In the domain of spoken word recognition, a hypothesis is a word w_i that the speaker might have produced, so the hypothesis space is an entire vocabulary of size N_w, and the observed data is the acoustic signal *A* perceived by the listener. Thus, an ideal observer model of spoken word recognition is given in Equation 2.3's restatement of Bayes' rule:

Equation 2.3

$$p(w_i|A) = \frac{p(A|w_i)p(w_i)}{\sum_{j=1}^{N_w} p(A|w_j)p(w_j)}$$

2.3.2. Prior Expectations in Spoken Word Recognition: Lexical Frequency

Equation 2.3 underlies the Shortlist B model presented by Norris and McQueen (2008). One of the most significant contributions of Shortlist B was a computational account of word frequency effects on spoken word recognition, a category of effects that ranks among the most robust findings throughout the psycholinguistic literature (e.g., Connine, Mullennix, Shernoff & Yellen, 1990; Dahan, Magnuson & Tanenhaus, 2001; Howes, 1954; Luce, 1986; Marslen-Wilson, 1987; Pollack, Rubenstein & Decker, 1960; Savin, 1963; Taft & Hambly, 1986). Following Norris' (2006) Bayesian Reader model of

visual word recognition, Shortlist B adopts each word's relative frequency as an estimate of its prior probability $p(w_i)$. Although this innovation is theoretically straightforward, it allowed Shortlist B to account for several classic effects, such as improved accuracy when subjects identify frequent words in noise compared to less frequent words (Luce & Pisoni, 1998).

Two observations about the role of the prior in a Bayesian model bear noting. First, a Bayesian spoken word recognizer will never "hallucinate" (*cf.* Norris et al, 2000) a word that bears no resemblance to the acoustic signal, no matter how frequent it may be. This follows from the fact that words that are entirely incompatible with some perceived signal are realized in the model with a likelihood $p(A|w_i) = 0$, and this will also entail that the posterior $p(w_i|A) = 0$. In the same vein, if no other words are consistent with a given acoustic signal, then even the rarest words can be clearly perceived (Doyle, 1890).

Second, even though a word's estimated frequency never changes in Shortlist B, the relative influence of the prior on subjects' behavior (as modeled by the posterior) will not be the same for all possible acoustic signals. Rather, the prior's influence on the posterior will be largest when the likelihood is most uncertain – that is, when there are many possible words that are somewhat consistent with the input. In contrast, when perceptual uncertainty is low, such that the likelihood is peaked over one or a small number of words, the same prior will be less influential on the posterior. As Norris and McQueen (2008) point out, this second observation matches findings of an interaction between word frequency and stimulus quality in word recognition accuracy data: the

more degraded a stimulus is by noise, the larger the observed advantage for frequent words (Luce & Pisoni, 1998).

It is clear that Shortlist B, and the Bayesian framework more generally, offers straightforward explanations for a broad range of phenomena spanning concepts as basic as lexical frequency, neighborhood density and lexical competition, perceptual confusability, and lexical influences on speech segmentation and word recognition. However, it is just as important to observe that these effects follow automatically from the basic principles that are mathematically required by a Bayesian model. As Norris and McQueen (2008) suggest, for many of the effects examined, their model could not be made to predict anything but the established finding and still be called "Bayesian." This stands in contrast to activation-based models of spoken word recognition, which require many architectural and dynamical assumptions and whose performance depend heavily on exact parameter settings within such models (Pitt, Kim, Navarro & Myung, 2006; see Norris, 2006 for discussion). That such a well-constrained model achieves such broad empirical coverage offers strong support for the notion that spoken word recognition might reflect optimal inference in the face of uncertain input.

2.3.3. Prior Expectations in Spoken Word Recognition: Sentential Context

Despite Shortlist B's successes, the rational analysis approach to computational modeling stresses the importance of revising a model's assumptions when a model cannot account for certain data. One type of data that is not accounted for by Shortlist B is the influence of sentential context on spoken word recognition. Indeed, the model assumes that the probability of each word in a sequence is independent of any other (non-

overlapping) words in a multi-word speech signal. Clearly, this assumption is not warranted.

In acknowledging this fact, Shortlist B's creators suggest how a more complete Bayesian model might approach this issue: "In all of the simulations reported here, we assume that [a word's prior probability] can be approximated by the word's frequency of occurrence in the language. However, [the word's prior] will also be influenced by factors outside the scope of the present model, such as semantic or syntactic context" (Norris & McQueen, 2008, p. 362). Since words do not occur randomly in language, an optimal listener's prior expectation over which words are likely should be highly context-dependent. Just as a model assuming that all words are equally likely fails to explain effects of word frequency, a model that assumes that some word is equally likely to occur in every context will necessarily fail to explain effects of sentential context.

The main goal of the present work is to relax the assumption of a context-independent prior. To do so, we begin with the basic approach of Norris and McQueen (2008), but – in the tradition of rational analysis – we update their assumptions in order to investigate whether behavioral patterns of context effects on spoken word recognition can also be explained by an ideal observer model. We also diverge from some other aspects of Shortlist B, most notably by adopting a model of words' likelihood functions that explicitly takes into account acoustic cues in the speech signal (see also Clayards, Tanenhaus, Aslin & Jacobs, 2008; Feldman et al, 2009; Feldman et al, 2013). This approach emphasizes the power of the Bayesian framework to explain lawful, fine-grained variability in how cues as disparate as sentential context and acoustic input interact during spoken word recognition.

2.4. *BIASES*: Bayesian Integration of Acoustic and Sentential Evidence in Speech

As already discussed, Bayes' rule describes the optimal way of combining two information sources – prior knowledge about which words a listener is likely to encounter, incorporated into the prior term $p(W)$, and acoustic data perceived by a listener, incorporated into the likelihood term $p(A|W)$.[2] However, it is also useful to invoke another common interpretation of Bayes' rule that is particularly applicable to modeling the effects of preceding context on spoken word recognition. As suggested by the standard nomenclature of *priors* and *posteriors*, Bayes' rule is often presented as an equation describing *optimal belief updating*. Put simply, if $p(W)$ indexes a listener's set of beliefs about how likely each possible word is *prior* to observing the relevant data (A), then $p(W|A)$ represents a listener's updated set of beliefs about the identity of the unknown word *after* integrating the new perceptual data, A.

Since the posterior, $p(W|A)$, depends on the prior, $p(W)$, and the likelihood, $p(A|W)$, it is intuitive that the likelihood drives the updating of a listener's beliefs. For instance, when the newly observed A is highly unlikely to be a token of a particular word (w_j), then the prior belief for w_j is revised downwards, rendering the posterior belief $p(w_j|A)$ smaller than the prior expectation $p(w_j)$. On the other hand, the more representative of w_j the observed signal A is, the more $p(w_j)$ will be revised upwards, causing $p(w_j|A)$ to gain support relative to other words. Within incremental sentence processing theories (Hale, 2001; Levy, 2008; Marslen-Wilson, 1973, 1975), this updating

[2] By convention, we use capital italicized letters to refer to a random variable. For instance, the prior distribution $p(W)$ defines the prior probability of each possible state of the random variable W. That is, if there are N_w words that a listener could hear, then $p(W)$ is a vector with N_w entries, such that each word w_j has some prior probability $0 \leq p(w_j) \leq 1$ and the sum $\sum_{j=1}^{N_w} p(w_j) = 1$.

process can be thought of as iterative, such that, after integrating each new piece of information or at each new time-step, the newly computed posterior becomes the updated prior for the next step in time.

Our model adopts this perspective in order to incorporate the influence of a preceding sentence context, *C*, on the recognition of a spoken word. Rather than assuming a static prior across all contexts, $p(W)$, we assume that listeners make use of a *conditional prior*, $p(W|C)$, such that their prior lexical expectations depend on the context up to that point (e.g., Altmann & Kamide, 1999; Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995; Kamide, Altmann & Haywood, 2003). Upon observing some subsequent speech, *A*, an ideal speech recognizer should update its contextually-conditioned prior beliefs by evaluating the probability that *A* was a token of each possible word. Given the simplifying assumption that listeners expect words to be pronounced with roughly the same acoustic form irrespective of which words preceded it (an assumption we address in greater depth in Chapter 4), the probability that *A* was a token of word w_i is given by Bayes' rule (Equation 2.4):

Equation 2.4

$$p(w_i|C,A) = \frac{p(w_i|C)p(A|w_i)}{\sum_{j=1}^{N_w} p(w_j|C)p(A|w_j)}$$

The model presented in Equation 2.4 serves as the basis for the remainder of this chapter. It represents a way of identifying which words were probably present in an imperfectly perceived speech signal by combining information from the preceding sentential context with subsequent acoustic cues. With this function in mind we will refer to this model as the *BIASES* model, short for *Bayesian Integration of Acoustic and Sentential Evidence in Speech*. As we will show, the model's name also foreshadows the

type of effect that it predicts should result when sentential evidence is brought to bear during word recognition. Next, we detail our implementations of the two fundamental components of BIASES: the context-dependent conditional prior, $p(W|C)$, and the likelihood function that relates words to their acoustic forms, $p(A|W)$.

2.4.1. Conditional Prior: A Model of Listeners' Contextual Knowledge

To define a conditional prior $p(W|C)$ for BIASES, every lexical candidate w_i must be assigned a probability of occurrence following each possible context. A conditional prior has two basic properties. First, in general, for a given context, some words will be more expected than others. This property is what makes any prior (conditional or not) informative: if all words are equally likely in some context, then the posterior is proportional to the likelihood alone. This is clearly not the case in human language, and listeners do clearly do not treat all words as being equally likely in a particular sentence context. Second, and in contrast to previous work (e.g., Norris & McQueen, 2008), different contexts will support the same word to different extents. It is this property that makes a prior conditional: $p(w_i|C = c_1)$ need not equal $p(w_i|C = c_2)$. As already discussed, whereas Shortlist B employed a context-independent lexical frequency prior, a key goal of BIASES is to incorporate a conditional prior that more accurately assumes listeners' access to context-dependent lexical expectations.

To do so in a computational model like BIASES, we must quantify the level of support that a given context provides for a word. It is undoubtedly the case that many factors collude to create a listener's expectation for any given word. A complete model of how context influences the probability of subsequent words would certainly depend on semantic (e.g., Borsky et al, 1998) and syntactic (e.g., Fox & Blumstein, in press)

information contained within the preceding linguistic context, but it would also depend on many other information sources that are available to a listener. For instance, a full model would need to address how listeners make pragmatic inferences about the implicatures in prior linguistic context (e.g., Rohde & Ettlinger, 2012), how listeners might employ speaker-specific and situation-specific knowledge about likely words or grammatical structures (e.g., Fine & Jaeger, 2013; Fine, Jaeger, Farmer & Qian, 2013; Kamide, 2012; Horton, 2007; van Berkum, van den Brink, Tesink, Kos & Hagoort, 2008), and how listeners treat knowledge about which words or concepts have recently been uttered in a discourse or are in common ground (e.g., Horton & Keysar, 1996), to name just a few. Quantifying the influence of such factors on subjects' lexical expectations is clearly not trivial, and doing is beyond the scope of the current modeling effort. Instead, we focus on an admittedly limited model of context in order to illustrate the explanatory power of the BIASES model, and the Bayesian framework more generally.

2.4.1.1. Conditional Expectations from *n*-gram Language Models

Most modern automatic speech recognition systems operate under the same fundamental hypothesis embraced by BIASES: that a word's context-independent frequency can capture only a fraction of the prior knowledge available during word recognition. The solution implemented in these models incorporates local semantic and syntactic context via language models (Jelinek, 1990, 1997). Put simply, an *n-gram language model* is a conditional probability distribution over lexical candidates given the *n*-1 immediately preceding words. As *n* increases, the probability distribution over possible words is conditioned on more information, and, consequently, the conditional

expectations for different lexical candidates become more fine-grained. For instance, a *bigram* language model $p(W_t|W_{t-1})$ estimates a word W_t's probability based on only the previous word, W_{t-1}, while a *trigram* language model $p(W_t|[W_{t-2}, W_{t-1}])$ estimates W_t's probability given that $[W_{t-2}, W_{t-1}]$ preceded W_t.[3] Intuitively, trigram language models make more specific predictions than bigram language models: fewer words are likely to follow ...*hated to*... than just *to*...

On the other hand, a *unigram* language model (n = 1) is simply a formal definition of a lexical frequency distribution. A word's frequency $p(w_i)$ can be computed by collapsing over all N_c possible preceding contexts via summation (referred to as *marginalization*; Equation 2.5).

Equation 2.5

$$p(w_i) = \sum_{j=1}^{N_c} p(w_i|C = c_j) = \sum_{j=1}^{N_c} p(W_t = w_i|W_{t-1} = w_j)$$

The observation presented in Equation 2.5 provides a mathematical justification for Norris and McQueen's (2008) original claim that (as they reiterated later) "frequency and context have the same explanation in a Bayesian model" (Norris, McQueen & Cutler, 2015, p. 4).

2.4.1.2. Consequences of Adopting an *n*-gram Language Model Prior

BIASES implements a language model as its conditional prior $p(W|C)$ for spoken word recognition. This decision has some obvious drawbacks, but also some important benefits. Under the strictest interpretation, the assumption entailed by employing an *n*-

[3] Note that trigram language models are order-sensitive. That is, in general, $p(W_t|[W_{t-2}, W_{t-1}]) \neq p(W_t|[W_{t-1}, W_{t-2}])$; intuitively, a listener's expectation for the word *pay* is not the same after hearing: *wanted to* and *to wanted*.

gram language model in this way is that all relevant information in *C* can be summarized by knowing the identities of the *n-1* words preceding the target word. Clearly, as discussed earlier, such a model is severely impoverished compared to listeners' actual contextual knowledge. That said, even a bigram language model constitutes far richer prior than Shortlist B's unigram language model (Norris & McQueen, 2008), which cannot account for any contextual effects on word recognition.

Language models might be considered among the simplest possible models capable of predicting context-specific modulation of word recognition. For example, a bigram language model would predict that, if w_i tends to follow w_{t-1} more often than w_j follows w_{t-1}, listeners should tend to identify an acoustic signal *A* that is perfectly ambiguous between w_i and w_j as w_i when the word preceding it was w_{t-1}. To the extent that such a model might account for some aspects of human behavior, it would suggest some commonalities between the detailed, linguistically relevant information contained within a listener's contextual knowledge and the transition probabilities between sequential words.

Note that it does *not* follow that listeners' models of context necessarily represent these word-by-word transition probabilities explicitly (see Levy, 2008 for discussion). This is another benefit of adopting a language model as BIASES's model of context for the purposes of the present computational-level analysis. The choice allows us to remain theory-neutral with respect to the actual representation of context used by listeners. We regard a language model as a convenient, useful tool to summarize some fraction of the information contained within a sentential context. As naïve as language models are, evidence suggests that they predict a number of measures in language processing, such as

reading times and eye fixations during reading (e.g., Hale, 2001, 2006; Levy, 2008; McDonald & Shillcock, 2003a, 2003b). By implementing a language model as the conditional prior in BIASES, we aim to test whether the predictive power of such models observed in psycholinguistic studies of reading will transfer to the domain of spoken language processing.

Of course, not all contextual influences on speech recognition will be explained by a standard language model. As just one example, Rohde and Ettlinger (2012) show that listeners' ratings of a phonetically ambiguous pronoun between *he* and *she* are biased towards the presumed gender of a referent who was the most likely "causer" of some event (*cf*. Garvey & Caramazza, 1974; McDonald & MacWhinney, 1995; Koornneef & van Berkum, 2006). Subjects preferentially rated phonetically ambiguous pronouns as *he* in sentences like *Noah frightened Claire because* [**?***e*] *drove 100 miles per hour*, but as *she* if the referents' names/genders were reversed. Such an effect would be difficult to explain with a basic language model, because the result appears to rely on inferential/causal reasoning above and beyond words' co-locational probabilities. Thus, the decision to use a language model as BIASES's model of context will prevent us from capturing effects like this one, but it is not implausible that Bayesian models of pragmatic reasoning (Bergen, Levy & Goodman, 2014; Frank & Goodman, 2012; Franke, 2009; Goodman & Stuhlmuller, 2013; Jager, 2012) could be incorporated into a Bayesian model of speech perception like BIASES.

While some sentential context effects are unlikely to find explanation in any sort of standard language model, the ability to account for other findings will depend on the precise specifications adopted for a language model. Indeed, most previously reported

semantic context effects could not be explained by a bigram language model. For example, the same word (*the*) immediately precedes the phonetically ambiguous target word in every stimulus sentence in Borsky and colleagues' (1998) study showing that subjects made more *GOAT* responses in *goat*-biased than in *coat*-biased sentences. A trigram language model, on the other hand, would likely explain at least some of the differences between *goat*-biased (*...milk the...*) and *coat*-biased (*...button the...*) sentences, but it would also predict that the entire semantic context effect observed in the study is driven by the word that appeared two words before the target, irrespective of the rest of the context (e.g., whether the sentence featured a *farmer* or a *tailor* as its subject). Whether or not this is true, what we hope to have made clear from the preceding examples is that the ability for a prior based on a language model to account for sentential influences on word recognition will depend on the sort of language model used and the sort of context effect examined.

A final observation about the consequences of selecting a language model as a prior regards a practical challenge it poses. Although more complex language models produce more fine-grained predictions, specificity of predictions trades off with sparsity of data (e.g., Katz, 1987). That is, as n increases, it becomes more difficult to estimate the probabilities associated with an n-gram language model because the number of possible contexts grows exponentially: if there are 10 words in a language, then there are 10 possible contexts in a bigram language model and $10^2 = 100$ two-word sequences for which to estimate probabilities. For a trigram language model in the same ten-word language, one must estimate all $10^3 = 1{,}000$ probabilities (each word in each of the $10^2 = 100$ possible two-word contexts). While most "possible" three-word sequences may

never occur in language, some sequences that are rare but do occasionally occur in language may never occur in a given corpus from which the language model is being estimated. If the corpus were to be taken as a perfectly reliable model of language, then those sequences will erroneously be assigned a prior probability of 0, making it impossible for a Bayesian word recognizer to observe that sequence in the future.

Although smoothing methods can be applied to ensure that all word sequences have some small but nonzero prior probability (Church & Gale, 1991; Dagan, Marcus & Markovitch, 1993; Good, 1953; Goodman, 2001; Katz, 1987; Jelinek & Mercer, 1985), these methods will also tend to reduce the model's ability to predict context-specific variability. When there is little information that might differentiate between the prior probabilities of two similarly rare sequences, they will tend to be treated as equally likely. Thus, although a bigram language model will lack a great deal of information that listeners will have access to, it will also provide a reliably-estimated language model capable of capturing context-dependent response patterns when those effects tend to be driven by the word immediately preceding the target that subjects are tasked with recognizing.

2.4.1.3. BIASES' Conditional Prior: A Bigram Language Model

With these factors in mind, and acknowledging the various limitations associated with language models, we adopted a bigram language model as the conditional prior for BIASES. As such, only the immediately preceding word influences the prior probability of the subsequent word. Furthermore, to examine how well BIASES could fit human behavior, the experiments we conducted utilized stimuli designed to elicit sentential context effects on word recognition that were driven by the immediately preceding word

(Fox & Blumstein, in press). In particular, the simulations and experiments reported here evaluated the influence of different function words (*to* vs. *the*) on the identification of the next word in a sentence. Fox and Blumstein showed that subjects were more likely to recognize a phonetically ambiguous word between *bay* and *pay* as *pay* when it was preceded by a sentence like *Brett hated to...* than when it was preceded by a sentence like *Valerie hated the...* According to BIASES, the basic explanation for this effect is that the two critical function word contexts (*to* vs. *the*) differentially influence a listener's prior expectations for immediately subsequent target words (*bay* vs. *pay*).

As we will show, despite this simplistic model of listeners' contextual knowledge, BIASES does remarkably well at predicting subjects' behavioral responses, including the overall pattern, patterns of subject-by-subject variability, and several other fine-grained quantitative predictions. We also conduct another set of simulations to examine the extent to which a richer model of context might account for additional, even more fine-grained context-specific patterns in our empirical results.

2.4.1.4. Additional Constraints on Prior Expectations: Forced-Choice

Our simulations invoke one other piece of contextual information that we assume subjects exploit. Since subjects receive a set of instructions before performing the experimental task in the laboratory, these instructions further constrain the conditional prior model of context that listeners use while recognizing words in the study. Specifically, we assume that, once instructed to identify the target word as either *bay* or *pay*, subjects assign all other words a prior probability of 0. This same assumption is almost universally implicit in other models of spoken word recognition. For instance, in TRACE (McClelland & Elman, 1986), responses during multiple-alternative forced-

choice tasks (such as phoneme or word identification experiments) are generated probabilistically from among a set of alternatives that is identified based on the task and stimuli (*cf*. Luce, 1959; McClelland & Rumelhart, 1981). By reading activation levels out from only a few "clamped" response alternatives, TRACE's decision model has the effect of nullifying any prior probability of a response from any other words outside of the predefined set. Similarly, the output nodes of other models (e.g., Shortlist: Norris, 1994; Merge: Norris et al, 2000; Shortlist B: Norris & McQueen, 2008) are pre-specified "on-the-fly" based on task demands. Although they are not strictly driven by theoretically interesting assumptions about human cognition, it makes sense that computational models of human behavior should account for such exogenous factors, and this is especially true for ideal observer models. As Norris (2006) puts it, which model will produce optimal behavioral responses "is critically dependent on the precise specification of the task or goal" (p. 330).

With this additional constraint, the summation over all possible words in the denominator of Equation 2.4 can be simplified to the sum of two terms: one proportional to the posterior probability of *pay* given *C* and *A*, and the other proportional to the posterior probability of *bay*. Equation 2.6 incorporates this assumption, giving the posterior probability of *pay*, where w_1 = *pay* and w_2 = *bay*.

Equation 2.6

$$p(w_1|C,A) = \frac{p(w_1|C)p(A|w_1)}{p(w_1|C)p(A|w_1) + p(w_2|C)p(A|w_2)}$$

This posterior probability distribution in Equation 2.6 gives the expected rate with which a subject should identify an acoustic stimulus as *pay* in a given context, if that subject were optimally combining the information sources we are assuming. To the

extent that subjects deviate from this behaviorally, Andersons's Principle of Rationality (1990) would demand that we update our assumptions. Note that the expected posterior probability of subjects making a *BAY* response, $p(w_2|C,A)$, is equal to $1 - p(w_1|C,A)$.

Finally, as pointed out by Feldman and colleagues (2009), in the case of modeling two-alternative forced choice, the posterior in Equation 2.6 can be rewritten to take the form of a logistic function. By dividing both the numerator and denominator of the right side of Equation 2.6 by the quantity in the numerator and applying inverse functions (exponential power and natural logarithm), Equation 2.6 can be rewritten as shown in Equation 2.7:

Equation 2.7

$$p(w_1|A,C) = \frac{1}{1 + e^{-\left[\log\frac{p(w_1|C)}{p(w_2|C)} + \log\frac{p(A|w_1)}{p(A|w_2)}\right]}}$$

2.4.1.5. Implementing BIASES' Prior: Corpus Estimates, Smoothing

An advantage of modeling subjects' responses in a two-alternative forced choice word identification task is that the implementation of BIASES' prior is quite flexible – flexible enough, in fact, that $p(w_1|C)$ and $p(w_2|C)$ need not actually be proper probabilities at all. To see this, one need simply note that the influence of the prior, $\log\frac{p(w_1|C)}{p(w_2|C)}$, is only dependent on the *ratio* of the prior probabilities of w_1 or w_2. A consequence of this is that their probabilities could just as easily be replaced by numeric values that are proportional to the words' relative prior probabilities. Because the prior in this implementation of BIASES is estimated from a bigram language model, counts from a corpus of the number of times w_1 and w_2 follow C in sequence will suffice ($\eta[C, w_1]$ and $\eta[C, w_2]$, respectively; see Equation 2.8).

Equation 2.8

$$\log\frac{p(w_1|C)}{p(w_2|C)} = \log\frac{p(W_t = w_1|W_{t-1} = C)}{p(W_t = w_2|W_{t-1} = C)} = \log\frac{\frac{\eta[C,w_1]}{\sum_{i=1}^{N_w}\eta[C,w_i]}}{\frac{\eta[C,w_2]}{\sum_{i=1}^{N_w}\eta[C,w_i]}} = \log\frac{\eta[C,w_1]}{\eta[C,w_2]}$$

Of course, many other words in the corpus besides w_1 and w_2 will also follow C (that is, $\sum_{i=1}^{N_w}\eta[C,w_i] \gg \eta[C,w_1] + \eta[C,w_2]$). However, the fact that the normalizing term (which represents the sum of all occurrences of C with any word) cancels out in Equation 2.8 reflects the assumption that subjects engaged in a two-alternative forced choice word identification task will only consider the relative contextual evidence for w_1 and w_2 in responding.

The data for BIASES's bigram language model were collected from the 2009 Google Books corpus (Michel et al, 2010). Rather than assuming that the corpus counts of the relevant bigrams (*the bay*, *the pay*, *to pay*, *to bay*) were perfect estimates for subjects' contextual knowledge, model-fitting (see Chapter 3) allowed for "add-alpha" smoothing (Lidstone, 1920). Under such a model, one value α is added to all bigram counts, and the fitting process selects the α that minimizes the overall deviation of the model predictions from the data. As mentioned earlier, while the benefit of smoothing is that it protects against overconfidence in our estimate of listeners' prior (especially in estimates of the probability of relatively uncommon bigrams), higher values of α tend to diminish the specificity of the predictions of the model. As the smoothing parameter α grows larger (and greater than the raw counts in the bigram language model itself), the model approaches a uniform distribution that renders w_1 and w_2 equally likely to follow every context.

2.4.2. Likelihood Term: Mapping an Acoustic Signal onto Lexical Forms

One of the most fundamental observations about speech communication is that there is no one-to-one mapping between words and their acoustic realizations (*cf.* Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). On one hand, it is clear that signal-to-word mapping is many-to-one: perfectly understandable productions of the same word can take on countless acoustic realizations that may differ from one another along many dimensions. On the other hand, signal-to-word mapping is also sometimes one-to-many: the same acoustic signal may, on different occasions, be perceived as different sounds (e.g., Ganong, 1980; Liberman, Harris, Hoffman & Griffith, 1957; Sawusch & Jusczyk, 1981), words (e.g., Borsky et al, 1998; Fox & Blumstein, in press), or sequences of words (e.g., Foss & Swinney, 1973; Kim, Stephens & Pitt, 2012). Together, these two facts underlie the likelihood term in BIASES and how it interacts with the model's conditional prior term.

2.4.2.1. Likelihood Functions: Many-to-One Mapping

The immediate function of the likelihood term $p(A|W)$ in BIASES is to formalize the many-to-one mapping from speech tokens to words. $p(A|W)$ is best described as a composite of N_w likelihood functions, where N_w is the number of words in the lexicon, because each word w_i has its own likelihood function $p(A|w_i)$. $p(A|w_i)$ defines a listener's phonetically-detailed knowledge about how w_i tends to be pronounced. Implicit in each word's likelihood function is the notion that not all productions of a word will be equally clear, and some realizations will be more typical than others. Work examining the influence of *category goodness* and *internal category structure* has shown that fine-grained acoustic properties in speech modulate perception, recognition and lexical access

(Andruski, Blumstein & Burton, 1994; Blumstein, Myers & Rissman, 2005; Kessinger & Blumstein, 2003; McMurray, Tanenhaus & Aslin, 2002, 2009; Miller, 1994; Miller & Volaitis, 1989; Pisoni & Tash, 1974; Volaitis & Miller, 1992). This internal category structure is modeled within $p(A|w_i)$ by making the more typical realizations of w_i more probable than less typical realizations.

The model assumes an acoustic space, $\mathcal{A}$, comprised of all possible speech waveforms, and each a_x in $\mathcal{A}$ is a point within that multidimensional acoustic space. Each word, w_i, occupies some subspace of $\mathcal{A}$ comprised of all possible pronunciations of w_i. The likelihood function of w_i, $p(A|w_i)$, assigns some probability to every possible a_x. For most values of a_x, it will effectively be the case that $p(a_x|w_i) = 0$; after all, although each word in a lexicon can be pronounced in many[4] ways, most possible waveforms will bear no similarity to some given w_i. However, among those speech tokens (values of a_x) that might plausibly be exemplars of w_i, the ones that most resemble w_i will be most probable according to $p(A|w_i)$. In this way, the role of the likelihood term of BIASES, $p(A|W)$, is to evaluate, for each lexical candidate w_i, how representative of w_i a perceived speech token a_x is.

Of course, just as the likelihood function of w_i will ensure that $p(a_x|w_i) = 0$ for most a_x, the overall effect of $p(A|W)$ is that, for a given acoustic signal a_x, $p(a_x|w_i) = 0$ for most words. As discussed earlier, when $p(a_x|w_i) = 0$ for all words but one, only one word will have a nonzero posterior probability, and there will be no question about the identity of the token a_x. Indeed, given the multiplicity of available bottom-up, acoustic

[4] Indeed, because at least some relevant dimensions of $\mathcal{A}$ are continuous (e.g., VOT, vowel duration, formant values), BIASES assumes that the sample space of possible pronunciations of any word is infinite. Thus, formally, $p(A|w_i)$ must be a probability density function.

cues that comprise the many dimensions of $\mathcal{A}$, it may very often be possible to distinguish words and speech sounds from one another (see, e.g., Nearey, 1990, 1997). However, there is not always such a consistent mapping from a given acoustic signal to one (and only one) word. Put simply, while the many-to-one mapping between speech tokens and words lies at the heart of each word's likelihood function, the computational challenge that a spoken word recognition system must overcome arises due to the one-to-many (or at least one-to-more-than-one) mappings between a signal and multiple possible lexical candidates. In such cases, the system must adjudicate among the various words that the perceived signal resembles to any degree.

2.4.2.2. Phonetic Ambiguity: One-to-Many Mapping

Under what circumstances would an optimal listener believe that, for more than one word, $p(a_x|w_i) > 0$? Feldman and colleagues (2009) identified several factors responsible for creating uncertainty in the mapping of a signal onto a single, best-matching word. Here, we classify these factors into two general categories. In short, the noisier the environment is and the more acoustically similar two words are, the more likely it is that the perceived signal a_x will be ambiguous between the two words.

One source of uncertainty is noise, which distorts the speaker's production of a word (s_x) and can cause the perceived acoustic signal (a_x) to be ambiguous between multiple words, even when s_x may not have been. For instance, as already discussed, early research in the phoneme restoration paradigm (Warren, 1970; Samuel, 1981, 1996) showed that masking a short segment of uninterrupted, natural speech with a cough or white noise could render the corrupted signal (/**il*/) consistent with any of several lexical candidates (e.g., *wheel*, *heel*, *peel*, *meal*) (Warren & Warren, 1970; Warren & Sherman,

1974). In general, adding noise to stimuli tends to "smear out" a word w_i's likelihood function: to accommodate more variability in the acoustic signal due to noise, more values of a_x will count as possible realizations of w_i. The result of this smearing is that some values of a_x may come to correspond to multiple possible words (Luce & Pisoni, 1998; Warren & Warren, 1970) or sounds (e.g., Cutler, Weber, Smits & Cooper, 2004; Miller & Nicely, 1955; Smits, Warner, McQueen & Cutler, 2003; Warner, Smits, McQueen & Cutler, 2005).

While the effect of noise is to increase the uncertainty of the speech signal after it is produced, a second source of uncertainty emerges naturally and is inescapable, even in a completely noiseless environment. Distinct words are sometimes characterized by overlapping likelihood functions; this occurs when the acoustic space corresponding to one word intersects with that of another word. A trivial example illustrating this fact is the case of homophony: if a speaker produces s_x = /*baɪ*/, s_x could correspond to several words (*buy*, *by*, or *bye*) because all three words share virtually identical spaces of possible pronunciations (but see, e.g., Gahl, 2008). The present work considers a less extreme example of how phonetic ambiguity may lead to lexical ambiguity. While homophony inevitably leads to lexical ambiguity, the phonetic ambiguity examined here arises when the likelihood functions of a pair of word-initial segments (/*b*/ and /*p*/) overlap in acoustic space.

The primary acoustic dimension on which /*b*/ and /*p*/ differ is voicing, with tokens of /*p*/ tending to be realized with longer voice-onset time (VOT) values than tokens of /*b*/ (Lisker & Abramson, 1964). Figure 2.1 displays two theoretical acoustic cue distributions over VOTs: one for word-initial /*b*/ and one for word-initial /*p*/. Although tokens of each

category can generally be distinguished on the basis of VOT alone, the two categories' distributions overlap such that tokens with some intermediate VOT values could plausibly be an exemplar of either /*b*/ or /*p*/. Although other acoustic dimensions of a spoken word also provide reliable cues that can distinguish /*b*/-initial tokens from /*p*/-initial tokens (e.g., Klatt, 1975; Lisker, 1986; Miller & Dexter, 1988; Repp, 1984; Stevens & Klatt, 1974; Summerfield, 1981), much work has shown that, holding other variables constant, listeners perceive segments with some VOT values as phonetically ambiguous between /*b*/ and /*p*/ (Clayards et al, 2008; Connine, Blasko & Wang, 1994; Connine, Titone & Wang, 1993; Fox & Blumstein, in press; Ganong, 1980; Liberman, Harris, Kinney & Lane, 1961; McMurray, Clayards, Tanenhaus & Aslin, 2008; McMurray et al, 2002, 2009; Miller & Dexter, 1988; Miller et al, 1984; Pisoni & Lazarus, 1974; Toscano & McMurray, 2012; Wood, 1976).

A consequence of this phonetic ambiguity for spoken word recognition is that an acoustic token /***ʔ****eɪ*/ with a phonetically ambiguous VOT could correspond to either *bay* or *pay* (Fox & Blumstein, in press). Thus, speech tokens whose initial consonants have intermediate VOT values exhibit a one-to-many mapping from acoustic signal to lexical forms, and, as illustrated in Figure 2.1, this one-to-many mapping can be modeled by assuming that *bay* and *pay* have overlapping likelihood functions.

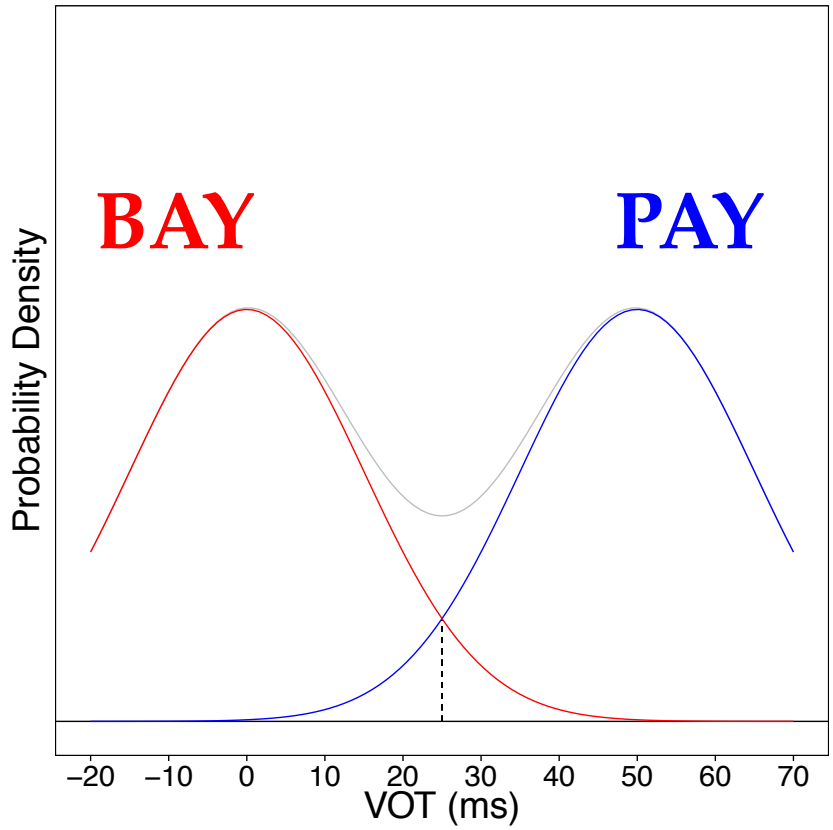

Figure 2.1. Examples of normally distributed probability density functions for two categories: /*b*/ and /*p*/, or for *bay* and *pay* under the assumption that these words are otherwise (i.e., besides VOT) identical in their acoustic cue distributions. The light grey line represents the marginal density function, showing the relative amount of total probability mass associated with each voice-onset time (VOT) across all categories. The dashed black line indicates the category boundary (χ), defined as the point in acoustic space (or the plane in acoustic space, if the likelihood model has more than one dimension) for which the probability density functions of two or more categories are equal. It can be equivalently defined as the point/plane for which the posterior probability distribution over a given set of categories is uniformly distributed when the prior probabilities of the set's members are also equal.

2.4.2.3. BIASES' Likelihood Term: A Mixture of Gaussians

An important issue that remains is the specification of each word's likelihood function. As already discussed, a complete model of the likelihood function for a word would define which acoustic signals could be recognized as a word, as well as how good an exemplar each possible signal would be. Although, in reality, such a model would be extremely complex and require a highly multidimensional space, the present model is far

simpler. Following prior work (e.g., Clayards et al, 2008), we assume that the likelihood functions of word-initial voicing minimal pairs (e.g., *bay* and *pay*) can be approximated by normal distributions over a single continuous dimension, VOT (see also Kleinschmidt & Jaeger, in prep; Kronrod, Coppess & Feldman, 2012; Munson, 2011). Under this assumption, listeners expect that if they were to perceive a given word w_i, the probability that it would be realized with different initial VOT values (A) is given by a normal distribution (see Equation 2.9A/B), where μ_i represents the mean initial VOT for w_i and σ_i^2 represents the variance in w_i's initial VOT.

Equation 2.9A

$$A|w_i \sim N(\mu_i, \sigma_i^2)$$

Equation 2.9B

$$p(A|w_i) = \frac{1}{\sqrt[2]{2\pi\sigma_i^2}} e^{-\frac{(A-\mu_i)^2}{2\sigma_i^2}}$$

If the acoustic form of w_i is assumed to be normally distributed, μ_i represents the most probable acoustic signal associated with w_i, and the further a stimulus is from this prototypical VOT, the less representative of w_i the exemplar will be (and the lower its likelihood will be). Since each word w_i has a Gaussian likelihood function $p(A|w_i)$ defined by its mean (μ_i) and variance (σ_i^2) parameters, the full likelihood model of BIASES, $p(A|W)$, which is a composite of all N_w words' likelihood functions, takes the form of a *mixture of Gaussians*. Gaussian mixture models are a common approach to statistical models of speech categorization and phoneme category learning (de Boer & Kuhl, 2003; Clayards et al, 2008; Dillon, Dunbar & Isardi, 2013; Feldman et al, 2009, 2013; McMurray, Aslin & Toscano, 2009; Toscano & McMurray, 2010; Vallabha,

McClelland, Pons, Werker & Amano, 2007) in which there exists some number of categories, and the exemplars of each category are normally distributed according to its category's mean and variance (or covariance matrix, in the case of multiple perceptual dimensions). Note that, for the present formalization of BIASES, the likelihood distribution over VOTs for each phonetic category (/*b*/ vs. /*p*/) is equivalent to the likelihood distribution over VOTs for each word (*bay* vs. *pay*) because (1) these two words are assumed not to differ on any other acoustic dimensions besides the VOT of their initial stop consonants, and (2) these words are assumed to be deterministically related to their associated phonetic categories. Admittedly, as we discuss later, it seems likely that neither of these assumptions is warranted; Chapters 3 and 4 illustrate some interesting and insightful implications of relaxing these assumptions. In any case, under these simplifying assumptions, in order to infer the most likely category label for a given exemplar a_x, a_x must be compared to each category's likelihood function (see Equation 2.9B).

Previous work suggests that listeners' perceptual identification behavior can be approximated by modeling words' likelihoods with a Gaussian mixture model over VOT values. In a study by Clayards and colleagues (2008) listeners identified acoustic stimuli along a VOT continuum between two words (e.g., *beach* and *peach*), and subjects' identification functions were consistent with an optimal Bayesian recognizer's behavior. A between-subject manipulation in their study provides further support for the modeling of words' likelihood functions using a mixture of Gaussians approach: two groups of subjects were exposed to different distributions of VOTs that implied either high overlap or less overlap in the words' likelihood functions (i.e., either higher values of σ_i^2 or lower

values of σ_i^2, respectively). Results showed that, when the overlap between the likelihood functions of *beach* and *peach* appeared to be higher, subjects perceived a greater number of the intermediate tokens from the continuum as ambiguous (Clayards et al, 2008).

Despite this modeling success, the Gaussian mixture model employed by Clayards and colleagues (2008) is not without its weaknesses. For instance, although they manipulated the acoustic variability of the stimuli for different subjects, the stimuli presented to any one group mimicked likelihood functions that had equal variance terms for both candidate words (e.g., *beach* and *peach*). Their model, in turn, also presumed that $\sigma_b^2 = \sigma_p^2$. In reality, it is not generally the case that the initial VOTs of words with an initial /*b*/ and words with an initial /*p*/ are distributed with equal variance (Lisker & Abramson, 1964; Kronrod, Coppess & Feldman, 2012). In fact, VOT distributions, as measured in spoken word and segment production experiments (e.g., Baese-Berk & Goldrick, 2009; Fox, Reilly & Blumstein, 2015), tend to exhibit non-Gaussian skew, and evidence suggests that the distribution of VOTs of word-initial voiced stops in English (especially /*b*/) is highly bimodal (Lisker & Abramson, 1964; see also Docherty, Watt, Llamas, Hall & Nycz, 2011).

Nonetheless, despite their divergence from speech production data, computational models of speech perception that have adopted the mixture of Gaussians approach and assumed equal variance across categories have achieved substantial success in capturing the overall patterns associated with category goodness and internal phonetic category structure (Clayards et al, 2008; Feldman et al, 2009; Kleinschmidt & Jaeger, 2015). Because of this fact, and because of the computational benefits associated with adopting this simplification (namely, the existence of a closed-form likelihood function), the

likelihood model implemented in BIASES was identical to the mixture of Gaussians employed by Clayards and colleagues (2008). In principle, any likelihood function could replace that of Clayards and colleagues (2008) in BIASES. For now, we simply acknowledge that BIASES could be enhanced with more detailed (and realistic) likelihood functions that incorporate more acoustic cues (see, e.g., Feldman et al, 2013) and/or less simplistic distributional assumptions (see, e.g., Kleinschmidt & Jaeger, in prep; Kronrod, Coppess & Feldman, 2012).

Critically, though, unlike Clayards and colleagues (2008), and unlike other Bayesian models of speech perception that have adopted similar models of the likelihood term (Feldman et al, 2009; Kleinschmidt & Jaeger, 2015), the key innovation in BIASES is the inclusion of a context-dependent conditional prior which is integrated with the likelihood function. The model proposed by Clayards and colleagues (2008) fits the basic shape of subjects' responses to isolated words despite their assumption of equal prior probabilities for *beach* and *peach* because their likelihood model captures fundamental properties of subjects' signal-to-word mapping. Here, we adopt this same likelihood function in formulating BIASES in order to leverage the successes of their isolated word recognition model, while extending it to incorporate sentential context effects.

2.4.2.4. Comparing the Likelihood Terms in BIASES and Shortlist B

Finally, it is worth pointing out another fundamental difference between BIASES and Shortlist B (Norris & McQueen, 2008). In addition to its assumption of a context-independent prior based on lexical frequency instead of the conditional prior embraced by BIASES, Shortlist B also differs from BIASES in the mathematical form of its likelihood term. Although the likelihood function adopted in BIASES is identical to that of Clayards

and colleagues (2008) and closely related to that of Feldman and colleagues (2009), Norris and McQueen's (2008) Shortlist B takes a very different approach. Rather than explicitly relating the acoustic properties of the speech signal to phonemes or words, Norris and McQueen (2008) avoided specifying a likelihood function that would directly relate to an acoustic signal. Instead of relying on assumptions about the distributions of acoustic cues and the ways in which these different cues covary, Shortlist B's likelihood model abstracts over all of the acoustic cues in speech to capture, broadly, the confusability of different words in the lexicon. Specifically, they assume that – whatever likelihood model subjects use – the same one that underlies subjects' performance in spoken word recognition tasks should also underlie their performance in lower-level perceptual tasks. Under this assumption, McQueen and Norris (2008) used perceptual confusion data from a gating task (Smits et al, 2003; Warner et al, 2005) to infer subjects' likelihood functions and taught Shortlist B, for each diphone in Dutch (e.g., /*ba*/), how likely subjects should be to perceive that diphone as itself or any other Dutch diphone (e.g., /*ba*/, /*pa*/, /*da*/, /*bɪ*/, ...).

The obvious advantage of Shortlist B's approach is that it can remain agnostic as to the computations that map an acoustic signal onto one or more words, so it can be applied to a large vocabulary without making many assumptions about pre-lexical representations or pre-lexical processing. However, the key question that motivated the development of BIASES was how bottom-up and top-down information sources are weighted and combined during spoken word recognition. In light of this question, the nature of the likelihood function and how it fits into the larger computational framework will play an important role in understanding the predictions of BIASES, as we discuss

later. Thus, unlike Shortlist B, BIASES is capable of making fine-grained predictions about how acoustic-phonetic properties of speech and context-specific lexical predictions jointly modulate subjects' recognition of spoken words.

2.4.3. Integrating Prior Context and Perceptual Input in BIASES

Substituting the likelihood function (Equation 2.9) into the model of subjects' posterior (Equation 2.7), and simplifying based on the stated assumption of equal variance in the VOT distributions ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) for *pay* ($w_1$) and *bay* ($w_2$), yields Equation 2.10 (*cf.* Feldman et al, 2009):

Equation 2.10

$$p(w_1|A, C) = \frac{1}{1 + e^{-\left[\log\frac{p(w_1|C)}{p(w_2|C)} - g(\chi - A)\right]}}$$

where

$$\chi = \frac{\mu_1 + \mu_2}{2} \text{ and } g = \frac{\mu_1 - \mu_2}{\sigma^2}$$

Equation 2.10 represents the optimal (Bayesian) posterior probability that a particular acoustic signal (A) following a particular sentence context (C) is an exemplar of the word *pay*, given the assumptions outlined above. Following a rational analysis approach (Anderson, 1990), Equation 2.10 can also be interpreted as an estimate of the optimal rate of *PAY* responses subjects should make when responding to different stimuli in different sentence contexts during a two-alternative forced choice word identification task.

Finally, in order to evaluate the extent to which the predictions of BIASES are consistent with actual subjects' behavior, it is possible to simulate responses from BIASES based on the assumption that, on a given trial (t) consisting of a context and

acoustic stimulus pairing (c_t, a_t), a subject's final identification decision (Z_t) is probabilistically generated from a Bernoulli distribution with $\theta_t = p(w_1|a_t, c_t)$ (see Equation 2.11).

Equation 2.11

$$Z_t|a_t, c_t \sim Bern\left(\frac{1}{1 + e^{-\left[\log\frac{p(w_1|c_t)}{p(w_2|c_t)} - g(\chi - a_t)\right]}}\right)$$

An oft-cited intuitive metaphor for Bernoulli-distributed random variables is the process of flipping a biased coin: the probability of a *PAY* response is like the probability of a coin-flip coming out heads, and different experimental conditions or stimuli affect the bias of the "coin" towards "heads" (θ_t) differently. In particular, Equation 2.10's posterior distribution describes the way stimulus and context conditions in a given trial influence the probability of a *PAY* response on that trial.

2.4.4. Conclusion and Next Steps

In sum, Equation 2.11 provides us with an explicit method of generating behavioral responses. The ability to simulate behavior from BIASES in this way affords many advantages. In Chapter 3, we take two somewhat different approaches to the simulation of behavioral data with BIASES. First, by providing to BIASES assumed values for all of the underlying parameters (μ_1, μ_2, σ^2, and $p(w_i|c_j)$ for every w_i and c_j relevant to the experiment) needed to generate behavioral responses, we can examine properties of the model's behavior, and examine the extent to which empirical data match those predictions. As we will demonstrate, this approach reveals that our chosen theoretical framework provides some much-needed clarity for research in the field of top-down effects, organizing a confusing literature rife with apparent inconsistencies.

Moreover, in the spirit of iterative updating of cognitive models that is fundamental to the rational analysis approach (Anderson, 1990), this approach allows us to identify enhancements to the model that are critical for capturing and "post-dicting" existing behavioral data. One major disadvantage to this approach is that, in order to generate responses, we must make even more assumptions about the latent structure of perceptual and cognitive processing that give rise to the behavioral responses we can observe.

On the other hand, a second approach allows us to *discover* features of the model underlying observed behavior, rather than making assumptions about the model's features. By generating data from BIASES over a very wide, weakly constrained range of possible parameter settings and comparing the simulated behavioral response patterns under different conditions to real data from human subjects, we can learn about the likely distribution of those parameters. As we will demonstrate, even though all we can actually observe is on the left side of Equation 2.11 (i.e., subjects' responses to different stimulus/context pairings), this approach allows us to infer the distributions of all of the unknown parameters that ultimately give rise to those responses. Moreover, this approach can be used to directly compare the relative fit and explanatory power of different theories and models that make disparate assumptions about aspects of auditory language processing.

As we will discuss, each approach has its own benefits and shortcomings, but both approaches can be leveraged to reveal important insights about the human speech perception system, and especially about how top-down information sources such as sentential context modulate word recognition.

Chapter 3

Exploring and Evaluating the BIASES Model of Spoken Word Recognition in Context

3.1. Understanding Top-Down Effects in BIASES

Although the theoretical and mathematical underpinnings of BIASES were presented in Chapter 2, Chapter 3 aims to explore the model's more fine-grained predictions about how higher-level (e.g., contextual) and lower-level (e.g., perceptual) information sources conspire during spoken word recognition to produce top-down effects on speech perception. To that end, Chapter 3 is divided into three main sections.

First, Chapter 3 examines the mathematical form of BIASES more closely. We implement BIASES and perform two preliminary simulation studies to illustrate how a minimalistic implementation of BIASES can replicate subjects' sensitivity to a preceding function word when identifying a stimulus that is phonetically ambiguous between a noun and verb (Fox & Blumstein, in press) and how the computational principles inherent to BIASES not only account for the overall pattern, but also provide fine-grained quantitative predictions about expected variability and asymmetries in the size of context effects on spoken word recognition.

Second, we consider a major problem that is often ignored in the literature, and especially by computational models: the enormous amount of unexplained variability in the size of top-down effects on speech processing. These issues are discussed with consideration for how these data can be captured by BIASES. The present model recasts previously overlooked or poorly understood behavioral patterns and asymmetries, suggesting that apparent inconsistencies in top-down effect on speech perception (*cf*. Pitt & Samuel, 1993) actually follow from the theoretical principles embodied by BIASES.

Illustrative simulations demonstrate the unique ability of BIASES to explain and predict lawful variability in the patterns of top-down effects across stimuli and across studies.

Finally, Experiment 3.1 is conducted in order to directly test one novel prediction made by BIASES, and the model's simulated behavior is compared to human performance on an auditory word identification task. These new experimental data are also used in two model comparison analyses to demonstrate that the utility of this computational model extends beyond providing a theoretical framework for contextual influences on word recognition. BIASES also represents a novel tool for comparing psycholinguistic theories about the two inputs to the model, including both the lower-level pre-lexical processing of speech that maps speech sounds to words, and the higher-level processing of sentences that reflects how listeners utilize contextual and linguistic information during auditory language processing.

Overall, the results of the simulations and the experimental analyses suggest that subjects' recognition of spoken words in context exhibit certain hallmarks of a Bayesian cue integration system. Generally speaking, BIASES highlights the fact that top-down effects on speech perception offer a unique window into perceptual processing, cognitive processing, and the interface of cognitive and perceptual representations in human language function.

3.1.1. Overview of the Mathematical Form of BIASES

Recall Equation 2.10's statement of the form of the posterior probability distribution, reproduced in Equation 3.1 (substituting in a new term, Π, to summarize the effect of the prior):

Equation 3.1

$$p(w_1|A,C) = \frac{1}{1 + e^{-[\Pi - g(\chi - A)]}}$$

where

$$\Pi = \log\frac{p(w_1|C)}{p(w_2|C)}, \chi = \frac{\mu_1 + \mu_2}{2} \text{ and } g = \frac{\mu_1 - \mu_2}{\sigma^2}$$

As described in Chapter 2, Equation 3.1 represents the present model's estimate of $p(w_1|A,C)$, the probability that a target stimulus was *pay*, given the voice-onset time (VOT) of its initial stop, *A*, given that it followed sentence context *C*, and given that the listener is performing a two-alternative forced choice word identification task with two possible candidates (w_1 = *pay* and w_2 = *bay*). The acoustic forms of *pay* and *bay* are modeled as having Gaussian distributions with means μ_1 and μ_2, respectively, and shared variance term σ^2 ($\sigma^2 = \sigma_1^2 = \sigma_2^2$). $p(W|C)$ is reflects the strength of the contextual bias towards one of the other candidate word, and it is estimated from a corpus. Finally, Equation 3.1's logistic form is a consequence of a key non-linguistic constraint: the use of a two-alternative forced choice task.

There are three key terms within the sigmoidal posterior (see Equation 3.1): (1) Π, the term summarizing the relative prior support for the candidate words, (2) *g*, the logistic function's *gain* term, and (3) χ, which denotes the VOT that is exactly halfway between the category means. Here, we discuss the interpretation of *g*, χ and Π in turn and explore their role predicting the distribution of top-down effects in spoken word recognition.

3.1.1.1. Components of BIASES: Phonetic Category Structure (*g*)

The primary effect of *g* is to control the slope of the logistic, with higher values of *g* indicating a sharper identification curve. As implied by the definition of *g* in Equation 3.1, greater separation between the means of the *pay* and *bay* ($\mu_1 - \mu_2$) and lower

variability (σ^2) in the expected distribution of productions of *pay* and *bay* are associated with steeper slopes. Simulation Study 3.1 illustrates the nature of *g*'s influence on the posterior probability function and on the size of sentential context effects on word recognition. The details behind these stimulations and their key conclusions are described in Box 3.1.

Figures 3.1 and 3.2 illustrate the tradeoff between category variance and category separation. The shape (i.e., steepness) of the resulting posterior sigmoids in Figure 3.2 is affected by changing either the distance between the means of the underlying normally distributed density functions in Figure 3.1 (left vs. right panels of the figures) or the underlying category variance of the normal probability density functions (or, equivalently, the standard deviations, as indicated in the top vs. bottom panels of Figures 3.1 and 3.2).

Intuitively, and as described in Chapter 2, the less overlap there is between two words' likelihood functions (Figure 3.1), the fewer acoustic values (here, VOTs) there will be that are ambiguous between *pay* and *bay* (Figure 3.2). Note that *g* is the term that differed between groups in the study by Clayards and colleagues (2008). By manipulating the apparent variance of the VOT distributions of /*b*/- and /*p*/-initial minimal pair words (e.g., *beach* and *peach*), Clayards and colleagues were tapping into the denominator of *g*.

Finally, note that, after hearing a sentence context *C*, all other terms in the posterior remain constant no matter what stimulus *A* is presented to the listener. In particular, the prior information (Π) does not influence the slope of the posterior distribution in BIASES (due to a conditional independence assumption; see Chapter 2), while *g* determines the overall shape of the posterior distribution over the acoustic space.

Box 3.1. Description of Simulation Study 3.1

Goal: Illustrate influence of two aspects of underlying phonetic category structure on posterior probability function and size of sentential context effects.

Design: 4 simulated phonetic category structures in a 2 × 2 design

Parameters of BIASES Manipulated: $\mu_1 - \mu_2 \in \{64,36\}$, $\sigma^2 \in \{15^2, 20^2\}$

Parameters of BIASES Held Constant: $\chi = 32$, $p(w_1|c_1) = 0.75$, $p(w_1|c_2) = 0.25$

Results displayed in: Figures 3.1-3.5, Table 3.1

Key conclusions:

1. BIASES' gain parameter (*g*), which characterizes the slope of the sigmoidal posterior (*cf*. Feldman et al, 2009), is the ratio of $\mu_1 - \mu_2$ (the distance between the means of the two words' distributions over VOTs) and σ^2 (the shared variance of each word's VOT distribution).
2. Because $\mu_1 - \mu_2$ and σ^2 are collinear (having opposite effects on *g*), they are not identifiable parameters when fitting a model that assumes equal category variance. The tradeoff between these features of BIASES' likelihood model can be visualized in the top-right and bottom-left simulated phonetic category structures in Figure 3.1, where distinct likelihood functions yield identical posteriors (Figure 3.2-3.3) with the same gain parameter (see Table 3.1). Consequently, model-fitting in Chapters 3-4 assumes values for $\mu_1 - \mu_2$ (from Lisker & Abramson, 1964; for a similar approach, see Kleinschmidt & Jeager, 2015) and fits σ^2.
3. Although the magnitude of the effective category boundary shift between two prior contexts ($\chi_{c_2} - \chi_{c_1}$) depends on *g*, the maximum expected effect size ($\Delta_{\max}$) is independent of it (see Table 3.1; Figure 3.4 vs. 3.3). However, a narrower range of VOTs exhibit top-down effects, so it would be more difficult, practically speaking, to detect a large top-down effect size if VOTs are sampled from the space.
4. For all 4 likelihoods examined in Simulation Study 3.1, the locus ($\hat{a}$) of the maximum expected effect size ($\Delta_{\max}$) was consistently collocated with the category boundary (χ) (Figure 3.4). Note, however, that χ was confounded with the midpoint of the effective category boundaries for the prior contexts (χ_{c_1}, χ_{c_2}). We discuss this point in Simulation Study 3.2 (see Box 3.2).
5. When measured for each prior context relative to a neutral baseline, the expected effect size for any given VOT is asymmetrical (in general); the locus of the maximum effect size is at the midpoint between χ and the prior context's effective category boundary (χ_{c_i}) (Figure 3.5).

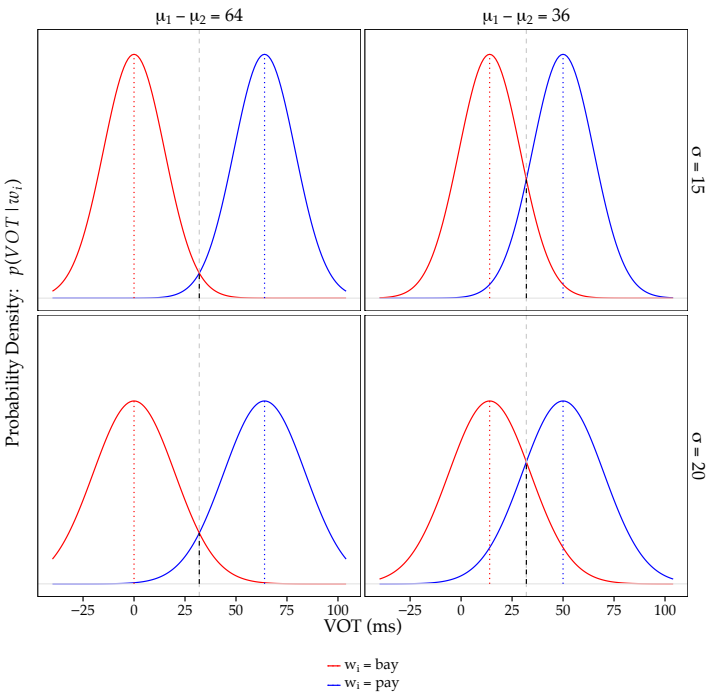

Figure 3.1. Results of Simulation Study 3.1: Influence of $\mu_1 - \mu_2$ and σ^2 on probability density functions, $p(VOT|w_i)$. $p(VOT|w_i)$: solid/colored curves; category boundary (χ): dashed/grey vertical line; μ_i for each $p(VOT|w_i)$: dotted/colored vertical lines

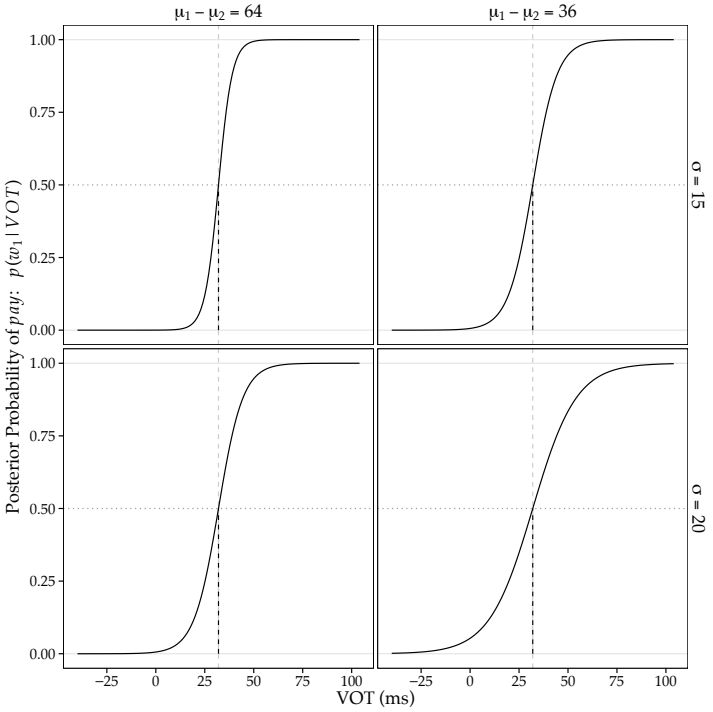

Figure 3.2. Results of Simulation Study 3.1: Influence of $\mu_1 - \mu_2$ and σ^2 on posterior probability function, $p(w_1|VOT, C = c_N)$: solid/black curves; χ: dashed/grey vertical line

3.1.1.2. Components of BIASES: Category Boundary (χ)

To understand the roles of the two remaining terms, Π and χ, consider, first, a hypothetical context c_N in which *w1* and *w2* are both equally probable *a priori*, such that $p(w_1|C = c_N) = p(w_2|C = c_N) = 0.5$ (*cf.* Clayards et al, 2008; Feldman et al, 2009). In this perfectly neutral context, $\Pi = \Pi_{c_N} = \log\frac{p(w_1|c_N)}{p(w_2|c_N)} = \log\frac{0.5}{0.5} = \log 1 = 0$. When this condition ($\Pi = 0$) is fulfilled, and when it also true that $A = \chi$, the entire right side of Equation 3.1 becomes $\frac{1}{1+e^0} = 0.5$. That is, χ is the point in acoustic space for which, if presented with that stimulus in a perfectly neutral context, a listener would, in theory, be equally likely to select either response: $p(w_1|A, c_N) = p(w_2|A, c_N) = 0.5$. This point, χ, is referred to as the *category boundary*. Note that, because BIASES assumes that the variability σ^2 associated with each word's likelihood function is equal, χ is guaranteed to be located exactly halfway between μ_1 and μ_2, as it is defined in Equation 3.1. However, note that if $\sigma_1^2 \neq \sigma_2^2 \neq \sigma^2$, then it is not, in general, true that $\chi = \frac{\mu_1+\mu_2}{2}$.

3.1.1.3. Components of BIASES: Prior Context (Π)

How, then, does the prior information contained within Π influence spoken word recognition? Because the influence of Π is constant for a given *C* (i.e., independent of *A*) the overall effect of the prior is to produce a translation (i.e., a horizontal shift) of the logistic function towards the mean of the less probable word (Feldman et al, 2009). Figure 3.3 (also from Simulation Study 3.1) illustrates this for the same distributions displayed in Figures 3.1 and 3.2. Shifting the logistic means that a stimulus with the same word-initial VOT will be more likely to be recognized as an exemplar of the

contextually-supported word (relative to context c_N, in which both lexical candidates are equally probable; black posterior distribution in Figures 3.2 and 3.3).

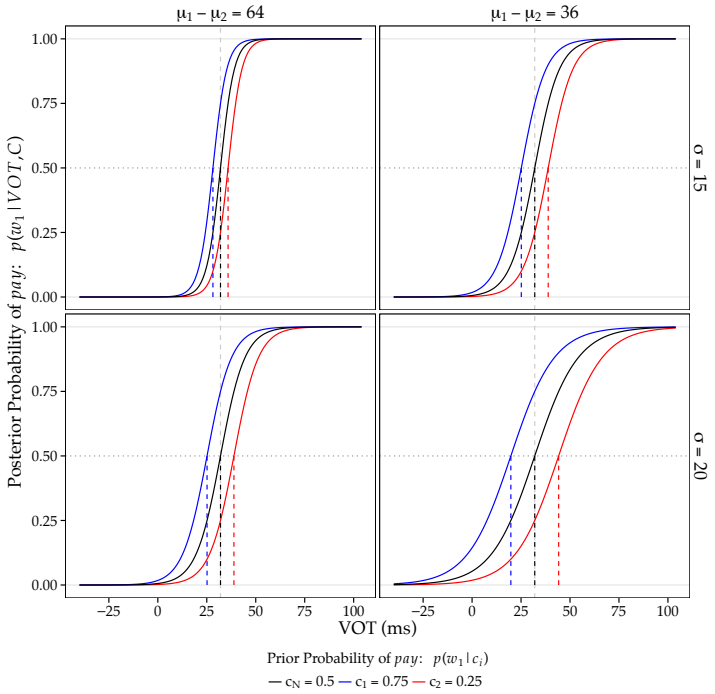

Figure 3.3. Results of Simulation Study 3.1: Influence of $\mu_1 - \mu_2$ and σ^2 on posterior probability function, incorporating prior contexts: $p(w_1|VOT, C = c_j)$: solid/colored curves; χ: dashed/grey vertical line; χ_{c_j} for each Π_C: dashed/colored vertical lines

In other words, Π induces a *bias* in the posterior sigmoid. Ultimately, this bias is realized for a given context C (for which $\Pi = \Pi_C$), in the movement of the location in acoustic space for which $p(w_1|A, C) = p(w_2|A, C) = 0.5$. We refer to this VOT value as the *effective category boundary* (χ_C). Unlike the "baseline" category boundary ($\chi = \chi_{c_N}$) defined earlier, which depends on characteristics of the likelihood function alone (*cf.* Clayards et al, 2008), the *effective* category boundary χ_C still, of course, depends on the likelihood function, but it is also context-dependent. Specifically, the magnitude and direction of its shift away from χ are given by $\chi_C - \chi = -\frac{\Pi_C}{g}$ (*cf.* Feldman et al, 2009). Figure 3.3 clearly depicts, for the same Π_C, the magnitude of the boundary shift should be

larger when g is smaller (i.e., shallower). Ultimately, the location of the effective category boundary for a given context is given by Equation 3.2:

Equation 3.2

$$\chi_C = \chi - \frac{\Pi_C}{g}$$

How can we summarize the meanings of each of the terms in BIASES (Equation 3.1)? Firstly, g encapsulates the model of the phonetic category structure that comprises the likelihood functions of *pay* and *bay* (Figure 3.1) and gives the posterior probability distribution its slope (Figure 3.2). While g controls the overall shape of the posterior, Π and χ convey information about the posterior's "location" in acoustic space. They determine, for instance, which VOTs will be most ambiguous. Along with g, χ also derives from the likelihood term, representing the location (in acoustic space) of the (unbiased) category boundary. Meanwhile, the influence of sentential context is completely contained within Π, which indexes the relative amount of support a context C provides to one or the other candidate word. It is Π's context-dependent *biasing* effect on the posterior distribution (Figure 3.3) that we focus on in the present model. Simulation Study 3.1 has already shown that the two elements of g ($\mu_1 - \mu_2$ and σ^2) have an influence on the magnitude of shifts in the effective category boundary, controlling for prior context. Next, we explore the basic predictions of BIASES, focusing on how different factors within the model influence the predicted size of the effect of prior context on subjects' word identification responses.

3.1.2. Towards Model-based Analyses of Top-Down Effects

3.1.2.1. Shifting of Invisible Category Boundaries

As discussed above, the "baseline" category boundary, χ, is the point at which a given VOT is equally likely to come from both categories' distributions (Figure 3.1). In the case considered here (and elsewhere, e.g., Clayards et al, 2008; Feldman et al, 2009), where the mixture distribution that comprises the model's likelihood term is effectively constrained to consider two Gaussian distributed categories with equal variance, χ lies at the midpoint of the two categories' means ($\frac{\mu_1+\mu_2}{2}$). χ is a fundamental property of phonetic category structure. In Simulation Study 3.1, all phonetic category structures assumed χ (=32), and this was held constant for all those examined.

However, in practice, χ is not straightforward to measure. For one, we cannot directly observe the phonetic category structure that underlies a subject's behavioral responses to acoustic tokens. Instead, we are bound to try to infer χ from a subject's identification or discrimination of speech tokens. However, even this is difficult, since the definition of χ presupposes equal prior probability. In reality, such a state is difficult to confidently tap into experimentally: due to pervasive effects of lexical and phonotactic frequency on subjects' recognition of speech sounds (e.g., Connine, Titone & Wang, 1993; Massaro & Cohen, 1983; Pitt & McQueen, 1998), many assumptions are required if one hopes to confidently infer its value (see Pitt & Samuel, 1993 for a discussion of this issue). In short, although the action of the prior in BIASES is attributed to a shift produced relative to an unobservable category boundary (see Equation 3.2), the same principles can be captured without explicitly relying on some assumed or inferred value of χ by examining relative effective category boundary shifts.

In other words, although the biased posterior distributions in Figure 3.3 (in blue and red) are, based on the cognitive model (Equation 3.1), computed by biasing the latent

(i.e., unobserved), neutral posterior (in black), in practice, an experimenter would compare data from two observed (biased) conditions to one another. Later, we do consider another type of experimental design that includes a designated "neutral" condition (e.g., Fox, 1984; Guediche et al, 2013; van Alphen & McQueen, 2001), and we argue that even in those conditions, subjects' responses are probably not completely unbiased, so the so-called neutral conditions are actually more likely to be just a third bias condition with an intermediate Π_C. In any case, for now, we focus on the far more common 2-condition experimental design.

Formally, to the extent that any shift in the effective category boundary, χ_{c_1}, is observed in context, c_1, it is relative to another context, c_2, with some other prior Π_{c_2}. Equation 3.3 is a generalization of Equation 3.2, giving the magnitude of the VOT boundary shift between two biased contexts.

Equation 3.3

$$\chi_{c_2} - \chi_{c_1} = \left(\chi - \frac{\Pi_{c_2}}{g}\right) - \left(\chi - \frac{\Pi_{c_1}}{g}\right) = \frac{\Pi_{c_1} - \Pi_{c_2}}{g} = \frac{\log\frac{p(w_1|c_1)}{p(w_2|c_1)} - \log\frac{p(w_1|c_2)}{p(w_2|c_2)}}{g}$$

Equations 3.3 supersedes Equation 3.2 (and is more general) because the neutral prior ($\Pi_{c_N} = \log\frac{0.5}{0.5} = 0$) that, by definition, characterizes $\chi_{c_N} = \chi$ could be substituted into Equation 3.3 to obtain Equation 3.2. Table 3.1 reports the relative effective category boundary shift for each simulation in Simulation Study 3.1, and they can be visualized as the difference between the VOT of the dashed/red line and that of the dashed/blue line in Figure 3.3. As previously mentioned, the boundary shift is inversely proportional to the value of g.

	$\mu_1 - \mu_2 = 64\ ms$	$\mu_1 - \mu_2 = 36\ ms$
$\sigma = 15\ ms$	$\chi = 32$	$\chi = 32$
	$\mathrm{g} = 0.28$	$\mathrm{g} = 0.16$
	$\hat{a} = 32$	$\hat{a} = 32$
	$\Delta_{\max} = 0.50$	$\Delta_{\max} = 0.50$
	$\chi_{c_2} - \chi_{c_1} = 7.72$	$\chi_{c_2} - \chi_{c_1} = 13.73$
$\sigma = 20\ ms$	$\chi = 32$	$\chi = 32$
	$\mathrm{g} = 0.16$	$\mathrm{g} = 0.09$
	$\hat{a} = 32$	$\hat{a} = 32$
	$\Delta_{\max} = 0.50$	$\Delta_{\max} = 0.50$
	$\chi_{c_2} - \chi_{c_1} = 13.73$	$\chi_{c_2} - \chi_{c_1} = 24.41$

Table 3.1. Summary of Results of Simulation Study 3.1: Influence of underlying phonetic category structure on posterior probability function and size of sentential context effects.

3.1.2.2. Boundary Shifts vs. Effect Sizes

However, even having skirted one practical issue by avoiding reliance on an unobservable parameter value, another practical issue remains that, ultimately, suggests that merely explaining top-down effects as arising from shifting category boundaries is not ideal for the goal of accurately assessing and predicting top-down effects in actual behavioral data.

To see this, consider the methodological and analytic techniques utilized by behavioral research examining top-down effects. Typically, experimenters construct an acoustic continuum and/or select a relatively small number of stimuli with discrete step sizes. In many cases, the durations of these step sizes are influenced by other practical constraints such as the need to splice waveforms at zero-crossings to avoid discontinuities which introduce acoustic artifacts such as clicks into the stimuli. Then, participants are presented with these tokens in different contexts for identification. The experimenter is effectively sampling from the subject's posterior distribution in order to characterize the listener's underlying prior and likelihood model. Finally, the

experimenter adopts some analytic technique, typically aimed at producing evidence that responses to the same acoustic tokens were categorized reliably differently between conditions (such as logistic regression or ANOVA over proportions of response-types by condition).

At no point in the process does the notion of a category boundary or some underlying horizontal shift arise. Of course, that does not mean it is not a useful characterization of the underlying model. It may suggest that either the "horizontal" shift of the biased sigmoid relative to another sigmoid (from another condition) or the "vertical" differences in the rate of categorization decisions is epiphenomenal, or even that both are. However, what is important about this observation for the present work is that the comparison of "effective category boundaries" across conditions is a theoretical construct and removed from real empirical research. That is, however fundamental or epiphenomenal a category boundary is to phonetic category structure, it is in some ways incidental to experimental research on top-down effects on spoken word recognition.

It should be noted that there are analysis techniques that are exceptions to the ones described above. For instance, some researchers do explicitly estimate boundaries for subjects in different conditions (e.g., Baum, 2001; Blumstein et al, 1994) and then compute statistics about the shift in the boundary between conditions. Obviously, this technique is neatly connected to the theory espoused here, but these statistics ultimately rely on estimates (not directly observed category boundaries). Statistics based on derived measures are necessarily less accurate than the original data themselves (see, e.g., Pitt & Samuel, 1993). Also, using a single summary statistic ignores many fine-grained details regarding the distribution of top-down effects (*cf*. Pitt & Samuel, 1993). We address this

analysis technique later (Chapter 4), ultimately showing that an explicit, model-based analysis approach allows for richer inferences about the nature of variability in top-down effects.

3.1.2.3. Predicting Effect Sizes

In order to better understand the way BIASES integrates prior information with bottom-up acoustic data with an eye towards the ultimate need to understand effect size (i.e., the "vertical" differences in response rates for a given acoustic stimulus), we first defined the function underlying BIASES' expected effect size when comparing the rate of *pay*-responses for any pair of contexts, for any VOT, as shown in Equation 3.4:

Equation 3.4

$$\Delta\left(\begin{bmatrix}\Pi_{c_1}\\ \Pi_{c_2}\end{bmatrix}, \begin{bmatrix}\chi\\ g\end{bmatrix}, A\right) = p(w_1|A, c_1) - p(w_1|A, c_2) = \frac{1}{1 + e^{-[\Pi_{c_1} - g(\chi - A)]}} - \frac{1}{1 + e^{-[\Pi_{c_2} - g(\chi - A)]}}$$

where

$$\Pi_{c_x} = \log\frac{p(w_1|c_x)}{p(w_2|c_x)} \qquad \chi = \frac{\mu_1 + \mu_2}{2} \qquad \text{and} \qquad g = \frac{\mu_1 - \mu_2}{\sigma^2}$$

For ease of exposition, we refer to the function defined in Equation 3.4 as $\Delta(A)$ with the understanding that $\Delta(A)$ is meaningless unless there are other parameter values provided to it ($\{\Pi_{c_1}, \Pi_{c_2}, \chi, g\}$). As expressed by Equation 3.4, $\Delta(A)$ is equal to the difference between the posterior probabilities of a *pay*-response to a stimulus (*A*) after context c_1 vs. c_2. As such, the function's shape will clearly depend on the same factors on which the posterior depends. For that reason, the first two arguments to the function are related to: (1) the biasing information in the prior of each posterior distribution (Π_{c_1} and Π_{c_2}), and (2) the two components of the posterior that are based on the phonetic category

structure defined by the likelihood function (χ and g, which do not change from context to context).

While it is simple to demonstrate that $\Delta(A)$ is the difference of two sigmoidal curves (more specifically, a sigmoid and that same sigmoid under translation), the function is not easy to express (*cf. difference of sigmoids membership function*: e.g., in Berkan & Trubatch, 1997). However, despite not having a simple, closed form, $\Delta(A)$ does have certain properties that are straightforward. More importantly, those properties are critical to the predictions and simulations discussed in this chapter and Chapter 4, so we review them now and illustrate several of them via simulation.[5]

Figures 3.4 and 3.5 are the final two Figures associated with Simulation Study 3.1 (see Box 3.1 for a summary of these simulations, and Table 3.1 for a summary of their results). Figure 3.4 illustrates the shape of $\Delta(A)$ over the acoustic space for same four simulated phonetic category structures in Figures 3.1-3.3. A few points bear special note:

Property 1. If $\Pi_{c_1} > \Pi_{c_2}$, then $\Delta(A) > 0$ for all values of A, although $\Delta(A)$ approaches 0 for values of A further from $\Delta(A)$'s peak, which we denote $\hat{a}$. Thus, in line with intuition, subjects should never show a reversal of a sentence context effect, on average.

Property 2. $\Delta(A)$ is symmetrical under the assumptions imposed on BIASES in Chapter 2; in particular, $\Delta(\hat{a} - x) = \Delta(\hat{a} + x)$ if $\sigma^2 = \sigma_1^2 = \sigma_2^2$.

[5] Note that $\Delta(A)$ is not a perfect tool for all purposes. For instance, although it does approximate the expected effect size after many trials in each context are completed, it is not meant to simulate behavioral data directly. After all, like the boundary shift, effect size is a derived measure that is epiphenomenal (from the explanatory standpoint of BIASES). Thus, the function is used for illustrative purposes, but all actual simulated behavioral data in Chapters 3 and 4 are generated using Equation 2.11 and subtracted to illustrate expected effect sizes.

Property 3. In general, it not possible to compute a definite integral for the sum or difference of two sigmoids. This is notable because the integral of $\Delta(A)$ (i.e., the total area between $\Delta(A)$ and the x-axis) is equal to the area between the 2 posteriors being compared: $p(w_1|A, c_1) - p(w_1|A, c_2)$. However, numerical methods can be used to approximate this value, and, conveniently, the total area under $\Delta(A)$, and therefore the total area between the $p(w_1|A, c_1)$ and $p(w_1|A, c_2)$ is equal to the magnitude of the effective category boundary shift between the two contexts (*cf*. Equation 3.3), as shown in Equation 3.5:

Equation 3.5

$$\int \Delta(A) = \int p(w_1|A, c_1) - \int p(w_1|A, c_2) = \frac{\log\frac{p(w_1|c_1)}{p(w_2|c_1)} - \log\frac{p(w_1|c_2)}{p(w_2|c_2)}}{g} = \chi_{c_2} - \chi_{c_1}$$

Property 4. If the maximum expected difference in the posteriors is located (in acoustic space) at $\hat{a}$ and has an effect size of magnitude $\Delta_{\max}$, Equations 3.6 and 3.7 give those values. $\hat{a}$ occurs at the midpoint of the two posterior probability distributions' effective category boundaries.

Equation 3.6

$$\hat{a} = \underset{a_x \in \mathcal{A}}{\operatorname{argmax}}\, \Delta(a_x) = \frac{\left(\chi - \frac{\Pi_{c_1}}{g}\right) + \left(\chi - \frac{\Pi_{c_2}}{g}\right)}{2} = \chi - \frac{\log\frac{p(w_1|c_1)}{p(w_2|c_1)} + \log\frac{p(w_1|c_2)}{p(w_2|c_2)}}{2g}$$

Equation 3.7

$$\Delta_{\max} = \Delta(\hat{a}) = \frac{1}{1 + e^{-\left[\frac{1}{2}\left(\log\frac{p(w_1|c_1)}{p(w_2|c_1)} - \log\frac{p(w_1|c_2)}{p(w_2|c_2)}\right)\right]}} - \frac{1}{1 + e^{-\left[\frac{1}{2}\left(\log\frac{p(w_1|c_2)}{p(w_2|c_2)} - \log\frac{p(w_1|c_1)}{p(w_2|c_1)}\right)\right]}}$$

Property 5. As is obvious from Equation 3.7, the value $\Delta_{\max}$ is independent of the specific characteristics of the phonetic category structure (i.e., the likelihood) in BIASES, including both χ and g. Unsurprisingly, $\Delta_{\max}$ *does* depend on the relative strengths of the biases of the prior contexts being compared. Simulation Study 3.2 further examines this issue (see Box 3.2; Figures 3.6-3.9; Tables 3.2-3.3).

One thing that we can conclude from these simulations and observations is that the effective category boundary shift – which underlies the model's explanation of top-down effects on spoken word recognition – is, indeed, closely tied to overall differences in the influence of two prior contexts on speech recognition (e.g., Property 3), but this shift does not tell the whole story of top-down effects on speech perception. According to BIASES, different effect sizes should be observed as a function of the underlying phonetic category structure (see Simulation Study 3.1) and as a function of the relative strengths of the biases of the two contexts being compared (see Simulation Study 3.2).

In short, BIASES predicts fine-grained variation in the size of top-down effects that should be observed in subjects' responses to different acoustic tokens in different sentential contexts. The distribution and shape of the $\Delta(A)$ curve (i.e., expected top-down effects as a function of VOT, for a given pair of contexts) depends on many factors. This general statement is of great theoretical interest because of the enormous variability and inconsistency in top-down effects observed in the literature (see, e.g., Pitt & Samuel, 1993). It is this issue to which we now turn.

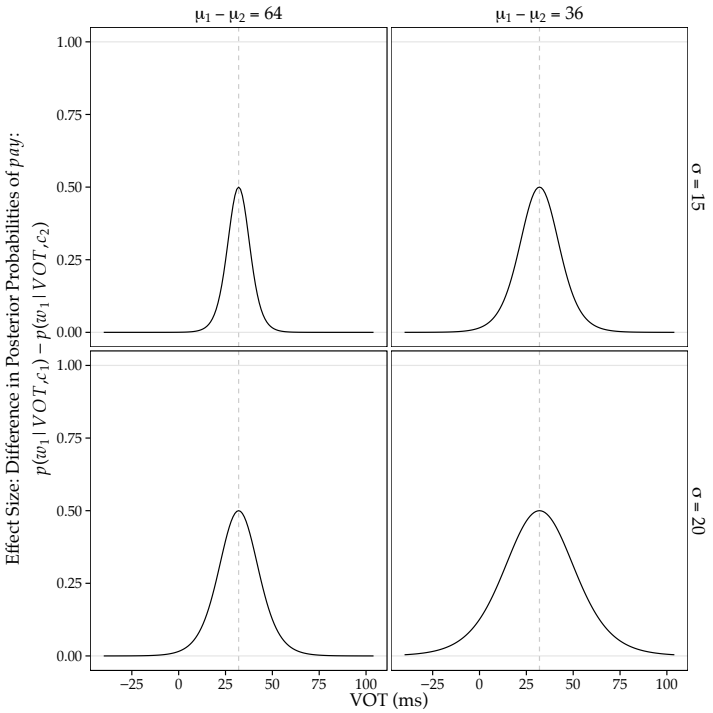

Figure 3.4. Results of Simulation Study 3.1: Influence of $\mu_1 - \mu_2$ and σ^2 on $\Delta(A) = p(w_1|VOT, C = c_1) - p(w_1|VOT, C = c_2)$. $\Delta(A)$: solid black curves; χ: dashed/grey vertical line.

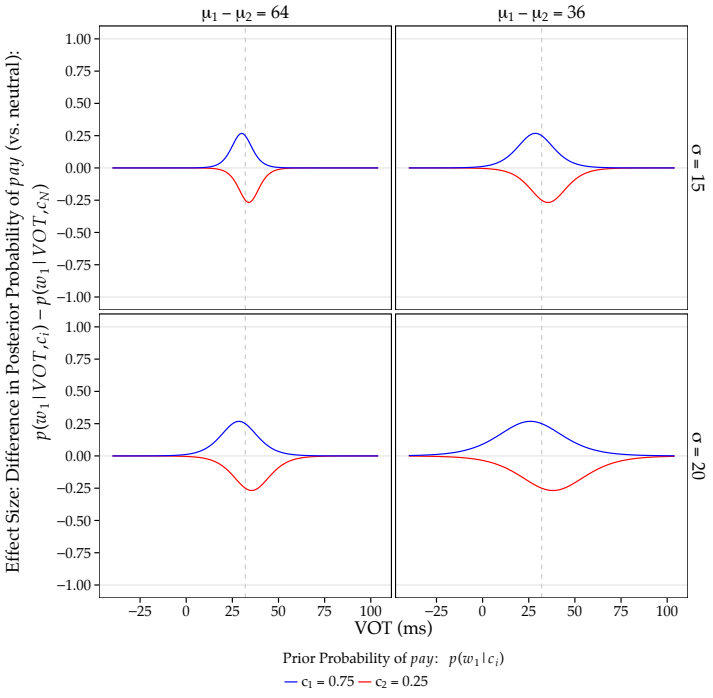

Figure 3.5. Results of Simulation Study 3.1: Influence of $\mu_1 - \mu_2$ and σ^2 on $\Delta(A) = p\left(w_1\middle|VOT, C = c_j\right) - p(w_1|VOT, C = c_N)$. $\Delta(A)$: solid/colored curves; χ: dashed/grey vertical line.

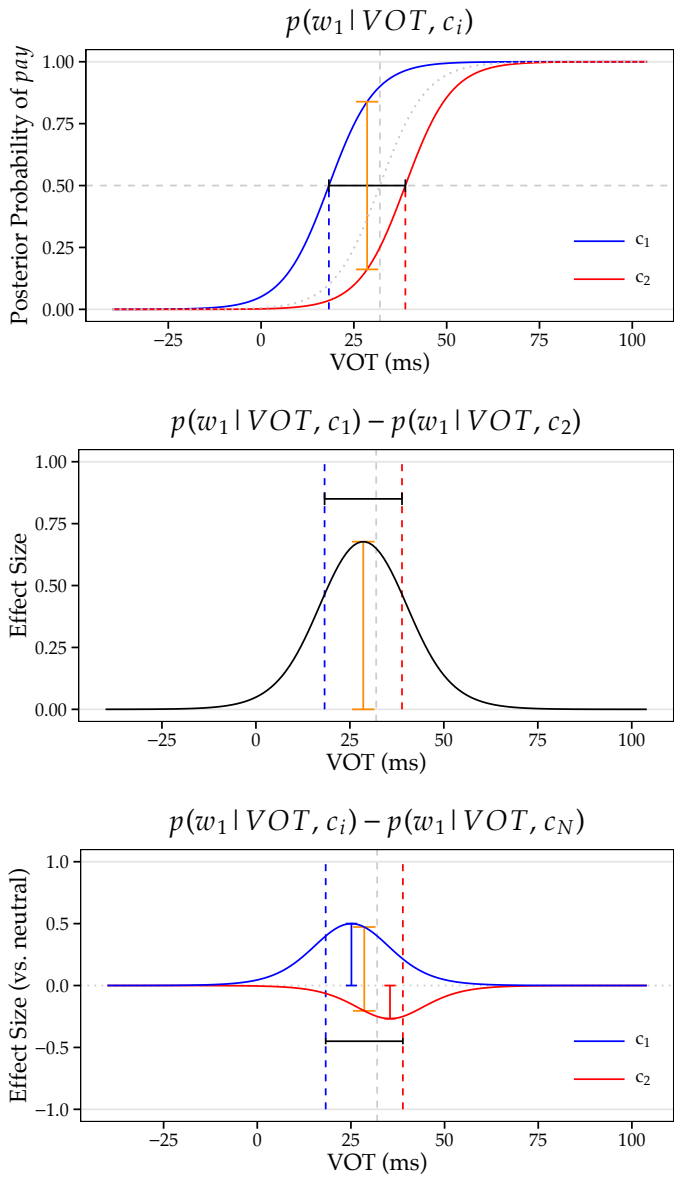

Figure 3.6. Example simulation from Simulation Study 3.2: Illustrates posterior probability distributions as a function of VOT and prior context ($p\left(w_1 \middle| VOT, C = c_j\right)$; top panel), effect size as a function of VOT for those two prior contexts ($\Delta(A)$; middle panel), and effect size as a function of VOT for each of those two prior contexts relative to a neutral baseline (bottom panel). All panels simulated in this illustration represent priors $p(w_1|VOT, c_1) = 0.9$ and $p(w_1|VOT, c_2) = 0.25$, with $\mu_1 - \mu_2 = 64$, $\chi = 32$, and $\sigma^2 = 20^2$. In all panels: χ: dashed/grey vertical line; χ_{c_j} for each Π_{c_j}: dashed/colored vertical lines; $\Delta_{\max}$: magnitude of solid/orange vertical marker; $\hat{a}$: VOT of solid/orange vertical marker; $\chi_{c_2} - \chi_{c_1}$: magnitude of solid/black horizontal marker. For top panel: $p\left(w_1 \middle| VOT, C = c_j\right)$: solid/colored curves; $p(w_1|VOT, C = c_N)$: dotted/grey curve. For middle panel: $\Delta(A) = p(w_1|VOT, c_1) - p(w_1|VOT, c_2)$: solid/black curve; For bottom panel: $\Delta(A) = p\left(w_1 \middle| VOT, C = c_j\right) - p(w_1|VOT, C = c_N)$: solid/colored curves; $\Delta_{\max}$ of each context's posterior relative to c_N: magnitude of solid/colored vertical markers.

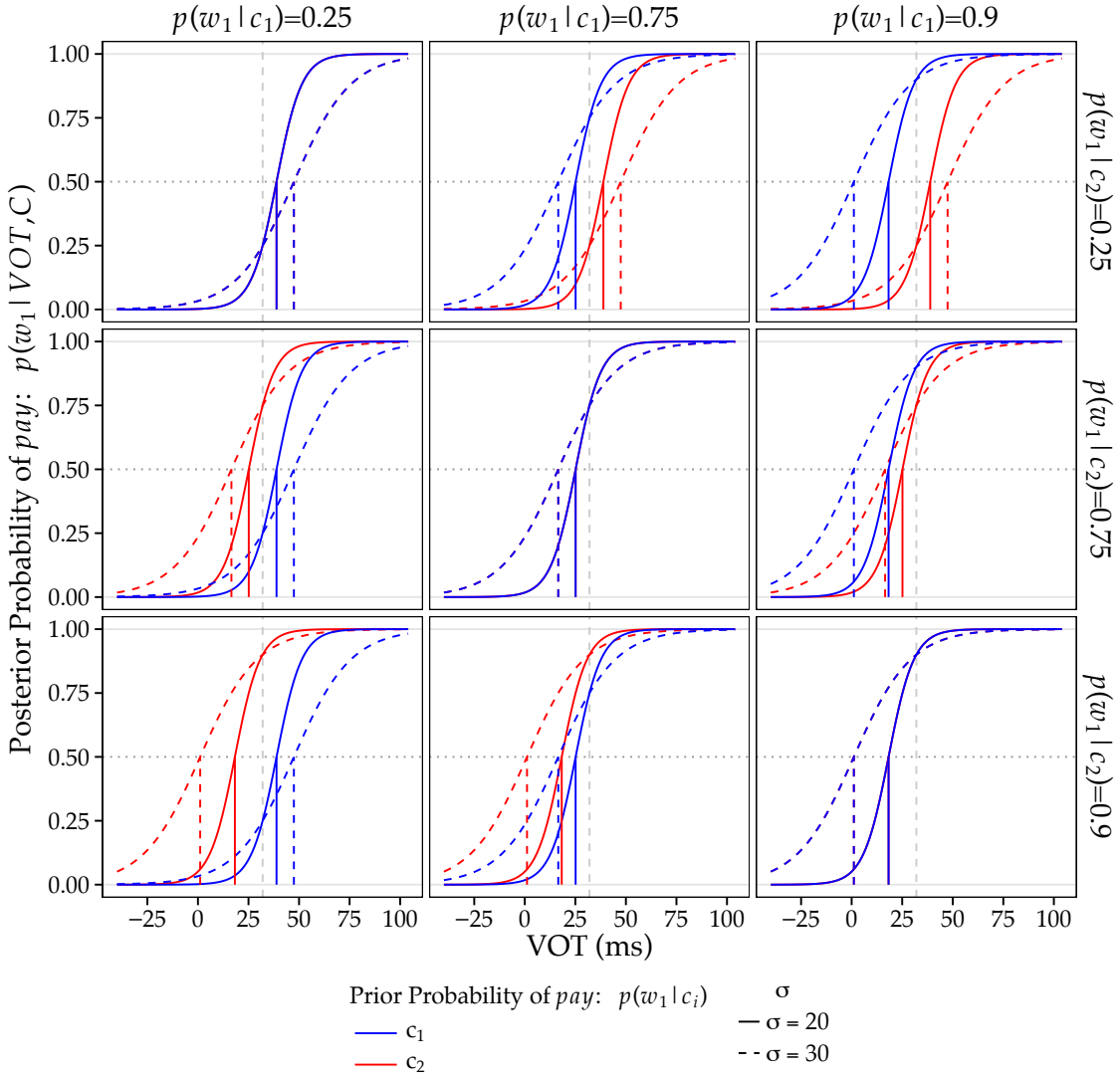

Figure 3.7. Results of Simulation Study 3.2: Influence of $p(w_1|C = c_j)$ and σ^2 on posterior probability function, incorporating prior contexts: $p(w_1|C = c_j)$: colored curves (solid: $\sigma = 20$; dashed: $\sigma = 30$); χ: dashed/grey vertical line; χ_{c_j} for each Π_C: colored vertical lines (solid: $\sigma = 20$; dashed: $\sigma = 30$)

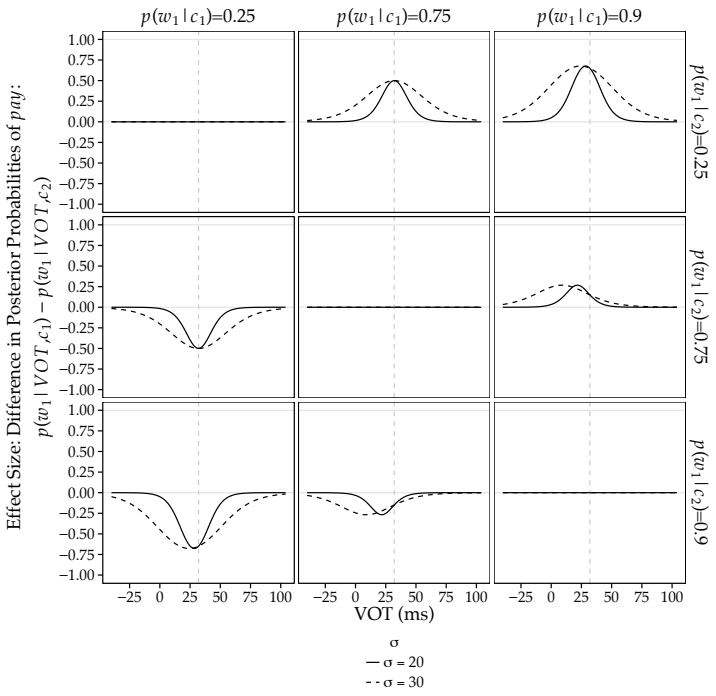

Figure 3.8. Results of Simulation Study 3.2: Influence of $p(w_1|C = c_j)$ and σ^2 on $\Delta(A) = p(w_1|VOT, C = c_1) - p(w_1|VOT, C = c_2)$; $\Delta(A)$: black curves (solid: $\sigma = 20$; dashed: $\sigma = 30$); χ: dashed/grey vertical line

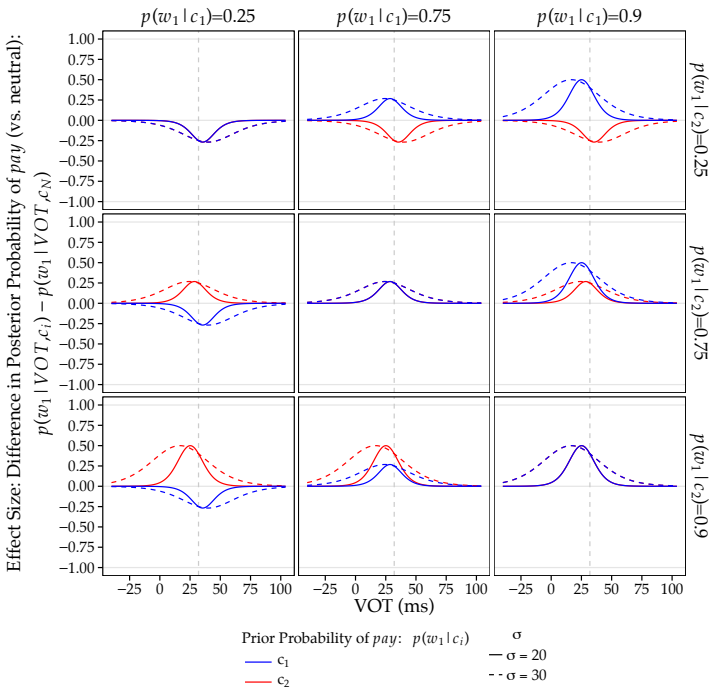

Figure 3.9. Results of Simulation Study 3.2: Influence of $p(w_1|C = c_j)$ and σ^2 on $\Delta(A) = p(w_1|VOT, C = c_j) - p(w_1|VOT, C = c_N)$; $\Delta(A)$: colored curves (solid: $\sigma = 20$; dashed: $\sigma = 30$); χ: dashed/grey vertical line

Box 3.2. Description of Simulation Study 3.2

Goal: Illustrate influence of the strength of contextual priors and one aspect of phonetic category structure on posterior probability function and size of sentential context effects.

Design: 2 phonetic category structures; 3 fully crossed levels of contextual bias in 2×3×3 design

Parameters of BIASES Manipulated: $\sigma^2 \in \{20^2, 30^2\}$, $p(w_1|C = c_1) \in \{0.25,0.75,0.90\}$, $p(w_1|C = c_2) \in \{0.25,0.75,0.90\}$

Parameters of BIASES Held Constant: $\chi = 32$, $\mu_1 - \mu_2 = 64$

Results displayed in: Figures 3.6-3.9, Table 3.2-3.3

Key conclusions:

1. Figure 3.6 illustrates the geometric interpretations of several critical variables in Chapter 3.
2. As in Simulation Study 3.1, the magnitude of the effective category boundary shift between two prior contexts ($\chi_{c_2} - \chi_{c_1}$) depends on g (Figure 3.7), but the maximum expected effect size ($\Delta_{\max}$) is independent of g. To see this, compare the solid and dashed curves in Figure 3.7. They peak at the same height. Note, also, that in Table 3.2 and Table 3.3, $\Delta_{\max}$ is the same for the same panel in each table. Table 3.2 lists the summary statistics for the simulations using $\sigma^2 = 20^2$ and Table 3.3 lists summary statistics for the simulations using $\sigma^2 = 30^2$. Each colored panel represent a pair of prior contexts, with tan panels showing no expected context effects (see Figures 3.7-3.8), blue panels having higher posteriors for c_1, red panels having higher posteriors for c_2, and darker panels (of each hue) corresponding to larger expected effect sizes. $\Delta_{\max}$ depends only on the prior contexts' biases' strengths.
3. Nonetheless, despite $\Delta_{\max}$ being independent of BIASES' likelihood function, the VOT at which the maximum expected effect size is found ($\hat{a}$) is not (see Figure 3.8). As Equations 3.6 and 3.7 suggest, $\hat{a}$ lies midway between the two priors' effective category boundaries, which depend on g. Note the divergence between Tables 3.2 and 3.3 in $\hat{a}$ for the same panel (i.e., priors).
4. Similarly, the magnitude of the effective category boundary shift between two prior contexts ($\chi_{c_2} - \chi_{c_1}$) depends on g and the priors (see Figure 3.7).
5. When measured for each prior context relative to a neutral baseline, the expected effect size for any given VOT is asymmetrical (in general); the locus of the maximum effect size is at the midpoint between χ and the prior context's effective category boundary (χ_{c_i}) (see Figures 3.6 and 3.9).

Tables 3.2-3.3. Summary of Results of Simulation Study 3.2: Influence of of $p(w_1|C = c_j)$ and σ^2 on posterior probability distribution and size of sentential context effects (Table 3.2: $\sigma^2 = 20^2$; Table 3.3: $\sigma^2 = 30^2$). Each colored panel represent a pair of prior contexts, with tan panels showing no expected context effects (see Figures 3.7-3.8), blue panels having higher posteriors for c_1, red panels having higher posteriors for c_2, and darker panels (of each hue) corresponding to larger expected effect sizes.

$\sigma^2 = 20^2$		$p(w_1\|VOT, c_1)$ = 0.25	$p(w_1\|VOT, c_1)$ = 0.75	$p(w_1\|VOT, c_1)$ = 0.90
$p(w_1\|VOT, c_2)$ = 0.25		$\chi = 32$ $g = 0.16$ $\hat{a} = 38.87$ $\Delta_{max} = 0.00$ $\chi_{c_2} - \chi_{c_1} = 0.00$	$\chi = 32$ $g = 0.16$ $\hat{a} = 32.00$ $\Delta_{max} = 0.50$ $\chi_{c_2} - \chi_{c_1} = 13.73$	$\chi = 32$ $g = 0.16$ $\hat{a} = 28.57$ $\Delta_{max} = 0.68$ $\chi_{c_2} - \chi_{c_1} = 20.60$
$p(w_1\|VOT, c_2)$ = 0.75		$\chi = 32$ $g = 0.16$ $\hat{a} = 32.00$ $\Delta_{max} = -0.50$ $\chi_{c_2} - \chi_{c_1} = -13.73$	$\chi = 32$ $g = 0.16$ $\hat{a} = 25.13$ $\Delta_{max} = 0.00$ $\chi_{c_2} - \chi_{c_1} = 0.00$	$\chi = 32$ $g = 0.16$ $\hat{a} = 21.70$ $\Delta_{max} = 0.27$ $\chi_{c_2} - \chi_{c_1} = 6.87$
$p(w_1\|VOT, c_2)$ = 0.90		$\chi = 32$ $g = 0.16$ $\hat{a} = 28.57$ $\Delta_{max} = -0.68$ $\chi_{c_2} - \chi_{c_1} = -20.60$	$\chi = 32$ $g = 0.16$ $\hat{a} = 21.70$ $\Delta_{max} = -0.27$ $\chi_{c_2} - \chi_{c_1} = -6.87$	$\chi = 32$ $g = 0.16$ $\hat{a} = 18.27$ $\Delta_{max} = 0.00$ $\chi_{c_2} - \chi_{c_1} = 0.00$

Table 3.2. Summary of Results of Simulation Study 3.2 (simulations utilizing $\sigma^2 = 20^2$).

$\sigma^2 = 30^2$		$p(w_1\|VOT, c_1)$ = 0.25	$p(w_1\|VOT, c_1)$ = 0.75	$p(w_1\|VOT, c_1)$ = 0.90
$p(w_1\|VOT, c_2)$ = 0.25		$\chi = 32$ $g = 0.07$ $\hat{a} = 47.45$ $\Delta_{max} = 0.00$ $\chi_{c_2} - \chi_{c_1} = 0.00$	$\chi = 32$ $g = 0.07$ $\hat{a} = 32.00$ $\Delta_{max} = 0.50$ $\chi_{c_2} - \chi_{c_1} = 30.90$	$\chi = 32$ $g = 0.07$ $\hat{a} = 24.28$ $\Delta_{max} = 0.68$ $\chi_{c_2} - \chi_{c_1} = 46.35$
$p(w_1\|VOT, c_2)$ = 0.75		$\chi = 32$ $g = 0.07$ $\hat{a} = 32.00$ $\Delta_{max} = -0.50$ $\chi_{c_2} - \chi_{c_1} = -30.90$	$\chi = 32$ $g = 0.07$ $\hat{a} = 16.55$ $\Delta_{max} = 0.00$ $\chi_{c_2} - \chi_{c_1} = 0.00$	$\chi = 32$ $g = 0.07$ $\hat{a} = 8.83$ $\Delta_{max} = 0.27$ $\chi_{c_2} - \chi_{c_1} = 15.45$
$p(w_1\|VOT, c_2)$ = 0.90		$\chi = 32$ $g = 0.07$ $\hat{a} = 24.28$ $\Delta_{max} = -0.68$ $\chi_{c_2} - \chi_{c_1} = -46.35$	$\chi = 32$ $g = 0.07$ $\hat{a} = 8.83$ $\Delta_{max} = -0.27$ $\chi_{c_2} - \chi_{c_1} = -15.45$	$\chi = 32$ $g = 0.07$ $\hat{a} = 1.10$ $\Delta_{max} = 0.00$ $\chi_{c_2} - \chi_{c_1} = 0.00$

Table 3.3. Summary of Results of Simulation Study 3.2 (simulations utilizing $\sigma^2 = 30^2$).

3.2. Evaluating BIASES

A central motivation behind the development of BIASES is to provide a theoretical explanation and computational framework within which to examine top-down effects from sentential context in spoken word recognition tasks. The foregoing work (Chapters 2 through 3.1) has focused on this task. However, BIASES provides more than a framework; as discussed above, BIASES also provides an explicit mathematical model that makes specific, fine-grained quantitative predictions about the distribution of top-down effects on spoken word recognition. Thus, it is important to evaluate the extent to which BIASES can account for observed variability in such effects, and the extent to which its novel predictions are borne out experimentally.

3.2.1. Observed Variability in the Size of Top-Down Context Effects

Despite the strong evidence for top-down effects on spoken word recognition (see Chapter 2), substantial heterogeneity remains to be explained in the fine-grained details of the results in studies of this class of phenomenon. Pitt and Samuel (1993) provide a thorough review of such variability for lexical effects, but we discuss a few examples here.

Observed effects vary depending on the source of bias (e.g., lexical, sentential, monetary payoff). Among lexical effects, the degree of top-down influence depends on the position of the manipulated phonetic cues in the word (e.g., word-initial: Ganong, 1980; word-medial: Connine, 1990; word-final: McQueen, 1991; see also Mattys, Melhorn & White, 2007). Among sentential context effects, the sizes of semantic, syntactic and pragmatic effects are not consistent. Even restricting analysis to top-down effects from syntactic sentential context on speech recognition, effect sizes vary greatly

(see Chapter 1; Fox & Blumstein, in press). Such effects are reminiscent of word frequency effects (Connine et al, 1993) in phoneme identification tasks (see Pitt & Samuel, 1993 for a related explanation of inconsistent lexical effects due to word familiarity/frequency).

In addition to varying with the nature of the biasing information, top-down effects also depend on characteristics of the acoustic stimuli that comprise the test continua. Most obviously and most consistently across studies, top-down effects are larger for more phonetically ambiguous stimuli and they vanish or are very small for tokens that are clearly identifiable. Other acoustic manipulations that reduce stimulus quality or otherwise render the tokens more ambiguous tend to be associated with larger effect sizes (e.g., Burton & Blumstein, 1995; McQueen, 1991; Pitt & Samuel, 1993). On the other hand, top-down effects are elusive when stimuli are more faithful to the phonetic properties of real speech and have a greater number of reliable bottom-up acoustic cues (e.g., Burton, Baum & Blumstein, 1989). Indeed, there is even some indication that the size and prevalence of top-down effects depends on the specific phonetic contrasts and the acoustic cues being manipulated in the stimuli (e.g., /*sh*/–/*ch*/ vs. /*sh*/–/*h*/ vs. /*sh*/–/*s*/ vs. /*b*/–/*m*/ vs. /*b*/–/*d*/ vs. /*b*/–/*p*/ vs. /*g*/–/*k*/ vs. /*t*/–/*d*/).

Furthermore, there is a high degree of individual subject variability in the extent to which subjects exhibit top-down effects, even within a homogenous population of healthy, monolingual English-speaking young adults with normal hearing (see, e.g., Chapter 1; Fox & Blumstein, in press). Far more variability exists when considering the size of such effects in elderly adults (e.g., Abada et al, 2008) or patients with aphasia (see Chapter 4; see also Baum, 2001; Blumstein et al, 1994; Boyczuk & Baum, 1999).

Finally, a review of the literature shows that there are also strong task effects and an influence of an experiment's demand characteristics on the observed size of top-down effects on speech recognition. Chapter 1 discussed the role of stimulus predictability, experimental task (phoneme vs. word identification), and response latency in determining the expected size of top-down effects (see Fox & Blumstein, in press; *cf*. Bicknell, Jeager & Tanenhaus, in press; Bicknell, Tanenhaus & Jeager, submitted; Connine, Blasko & Hall, 1991; McClelland, 1987; Pitt & Samuel, 1993; Szostak & Pitt, 2013; van Alphen & McQueen, 2001). Pitt and Samuel (1993) also acknowledge apparent modulations of top-down effects in mixed vs. blocked designs, and they highlight the potential for differences in measured effect sizes due to differences in the analytic techniques experimenters select.

In some cases, such variability may be due to chance. However, it is also possible that the observed asymmetries and inconsistencies are not merely noise, but are, in fact, systematic variation attributable to the basic principles underlying speech perception and the probabilistic Bayesian framework within which we are have formulated an explanation of sentential context effects on spoken word recognition. Because BIASES offers a formal, mathematical model of context effects on speech perception, it is possible to evaluate the quantitative predictions of the model in light of available data. In this way, not only can we validate many of the fundamental principles underlying BIASES, but we can also identify shortcomings of BIASES and take measures to improve the model's empirical coverage.

Next, we consider four sources of variability, examining whether and/or how BIASES might capture the observed irregularities, and, in the process, relaxing some of

the simplifying assumptions adopted when the model was introduced in Chapter 2. Specifically, we examine two sources associated with the likelihood term of BIASES (variability in the ambiguity of phonetic cues based on VOT and based on additional cues) and two sources associated with the prior term of BIASES (variability based in the strength of prior context and based on a "neutral" context).

3.2.2. Variability in the Ambiguity of Phonetic Cues: VOT

One well-documented source of variability in top-down effects on spoken word recognition is variability along a continuum; top-down effects are not typically observed for phonetically unambiguous endpoint stimuli. For example, stimuli with very short VOTs are not good exemplars of /*p*/, as reflected in the likelihood distribution for /*b*/ and /*p*/ in Figure 3.1. There is a vanishingly small probability that a word-initial /*p*/ will be pronounced with a VOT of 10 ms, so the posterior probability (see Figure 3.2) of a /*p*/-response for such a token is virtually zero. Even when the prior context strongly supports a word beginning with /*p*/ (see Figure 3.3) subjects are not likely to make a /*p*/-response; after all, the posterior in Bayes' rule is proportional to the product of the prior and the likelihood, so acoustic tokens that are not at all representative of /*p*/ (i.e., have a likelihood close to zero) will not tend to show reliable context effects. The same is true of stop consonant tokens with VOTs of 50 ms, for example, because the likelihood that that VOT is a token of *bay* is practically zero. This can be seen clearly in Figure 3.4, where no context effects are observed for these VOT values.

This pattern has been replicated in the literature, going back to Ganong's original lexical effect (1980); for instance, a large effect for intermediate VOTs and much smaller or nonexistent effects for endpoint VOT tokens can be seen in the data from Chapter 1

(Fox & Blumstein, in press; see Chapter 2 for discussion). This is not a strictly Bayesian pattern: many models are capable of capturing this sort of effect. However, the pattern exemplified in Figure 3.4 is a fundamental property of the Bayesian framework (e.g., Massaro, 1989; Norris & McQueen, 2008), rather than the consequence of a design choice within the model. This contrasts with other models, such as Merge (Norris et al, 2000), which prevents top-down information from influencing responses to endpoint tokens by implementing a "bottom-up priority rule" that only allows higher-level sources to affect decisions if the acoustic information is ambiguous (Norris et al, 2000). Importantly, in order to explicitly implement something like Merge's bottom-up priority rule, a model must define some additional computational machinery and/or assumptions to govern when bottom-up decisions are protected from contextual influences vs. when top-down information is integrated.

Although other cue integration models (e.g., Toscano & McMurray, 2010), which focus on the integration of multiple acoustic cues in speech perception, also exhibit *reliability-based cue-weighting* like the present model, BIASES also makes fine-grained quantitative predictions about the distribution of top-down effects across VOTs that can be compared to patterns in empirical data. To the extent that these specific predictions are borne out, it would suggest that there exist certain hallmarks of Bayesian cue integration in behavioral results. This issue is examined further later in this chapter.

3.2.3. Variability in the Ambiguity of Phonetic Cues: Additional Cues

Although, up to now, we have assumed that VOT is the primary acoustic dimension on which voiced and voiceless stop consonants (e.g., /*b*/ and /*p*/) are distinguished in speech perception (Liberman et al, 1961), listeners also make use of a

large variety of other acoustic cues in their judgments about voicing in natural speech stimuli (see, e.g., Klatt, 1975; Lisker, 1986; Miller & Dexter, 1988; Repp, 1984; Stevens & Klatt, 1974; Summerfield, 1981). A complete model of top-down effects on spoken word recognition, then, would include all of these cues in the likelihood model that maps acoustic stimuli to words.

Burton, Baum and Blumstein (1989) investigated one cue in particular – the amplitude of the burst of the stop consonant. In natural speech, VOT and burst amplitude co-vary (Lisker & Abramson, 1964; Pickett, 1980; Zue, 1976), even though most VOT continua hold burst amplitude constant in an attempt to isolate the effect of VOT duration on speech recognition. Burton and colleagues (1989) showed that not only were subjects sensitive to manipulations of burst amplitude as a cue to voicing in stop consonants, but that the size of top-down effects as determined by the emergence of a lexical effect in their responses differed depending on whether the stimuli in the test VOT continuum varied from token to token in both the burst amplitude *and* VOT or just in VOT. Top-down effects occurred when only VOT varied, and were not, in fact, significant when burst amplitude and VOT co-varied along the continuum (as in natural speech).

How might this asymmetry be explained within the Bayesian framework? Clearly, BIASES is not equipped to explain this effect under its original assumptions because it assumes that VOT is the only relevant acoustic cue to a word onset's identity. A simple adaptation, however, can explain how this effect emerges. First, we must incorporate both the burst amplitude and the VOT of a given stimulus into BIASES' likelihood function, $p(A|W) = p(burst, VOT|W)$. With this change, the likelihood function is two-dimensional instead of one-dimensional; note that, while this is still surely an

oversimplification, since many other acoustic cues also influence word recognition, it illustrates the adaptability of BIASES. Next, we created an arbitrary range of burst amplitudes that was higher for voiceless tokens than voiced tokens (*cf.* Lisker & Abramson, 1964).

Finally, a simulation was conducted to compare the expected size of top-down effects in responses to stimuli from two simulated VOT continua: a VOT continuum with a single burst amplitude across all token and a VOT continuum with VOT values that co-varied with burst amplitude. Arbitrary mean and variance parameters were selected from among those used in Simulation Study 3.1 (any choice shows the same basic pattern, but the actual size of the effective category boundary shift depends on this choice; see Figure 3.3 and Table 3.1). Figure 3.10 shows the posterior probability distributions for the two simulated continua in two biasing contexts (blue vs. red) and a baseline neutral context (black).

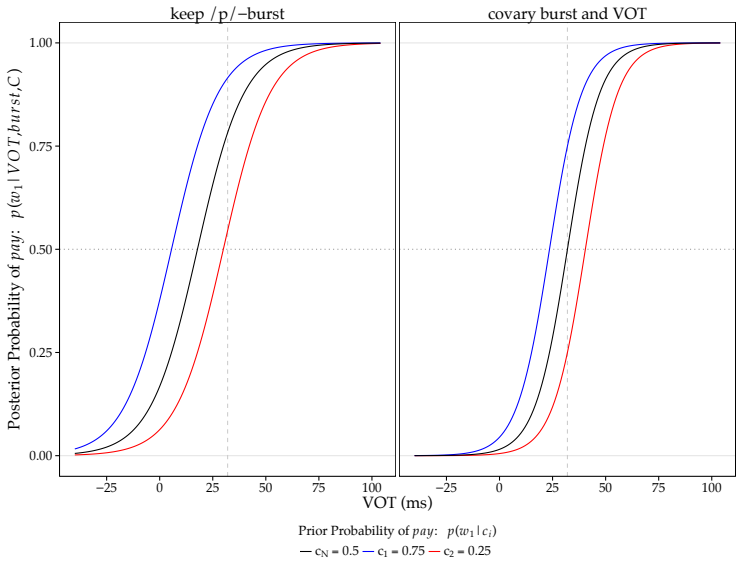

Figure 3.10. Results of two model simulations of posterior probability distributions assuming the same biasing/neutral contexts and identical underlying likelihood models. Simulations on the left and right only varied in whether of not the VOTs of the stimuli were correlated with burst amplitude of the simulated stimuli (right) or not (left).

As can be seen, the expected category boundary shift (and effect sizes) are smaller for the second simulated stimulus set (burst amplitude and VOT co-varied) than in the first simulated stimulus set (VOT varies with a constant value for burst amplitude: the mean amplitude of the voiceless stop's burst). It is important to note that the model is the same in both simulations (it is BIASES with the updated likelihood model to include burst amplitude as an acoustic cue) – only the characteristics of the simulated stimulus sets vary between the two simulations.

Next we consider two sources of variability related to the prior of BIASES.

3.2.4. Variability in the Strength of Prior Cues

Figure 3.11 is a reproduction of Figure 1.2 (in Chapter 1; Fox & Blumstein, in press), which shows the results of Experiment 1 from that study: the proportion of /*p*/-responses made to ambiguous tokens (i.e., the intermediate VOT values as defined in Chapter 1) from the *bay–pay* and *buy–pie* continua following noun-biasing (e.g., *Valerie hated the...*) and verb-biasing (e.g., *Brett hated to...*) sentence contexts. Recall that the *bay–pay* continuum was designed to be a *noun–verb* continuum and the *buy–pie* continuum was designed to be a *verb–noun* continuum. As explained in Chapter 1 and as can be seen in Figure 3.11, consistent with predictions, subjects exhibited a significant CONTEXT x CONTINUUM interaction, wherein they were more likely to make /p/-responses when the most common grammatical category of the /*p*/-endpoint was consistent with the grammatical cue provided by the preceding function word (*to* vs. *the*).

While this effect was quite robust, with the simple effects of CONTEXT being significant and in opposite directions in each level of CONTINUUM, the effect sizes were not identical. This can be seen clearly in Figure 3.11: the magnitude of the effect of

CONTEXT in the *buy–pie* continuum (β = 0.95) is smaller than the effect's magnitude in the *bay–pay* continuum (β = -1.37). Moreover, visual inspection of Figure 3.11 suggests that the primary level of CONTEXT that is driving the interaction is the verb-biasing (*to*) level: the proportion of /*p*/-responses is far more disparate between the two continua following the verb-biasing contexts than the noun-biasing contexts.

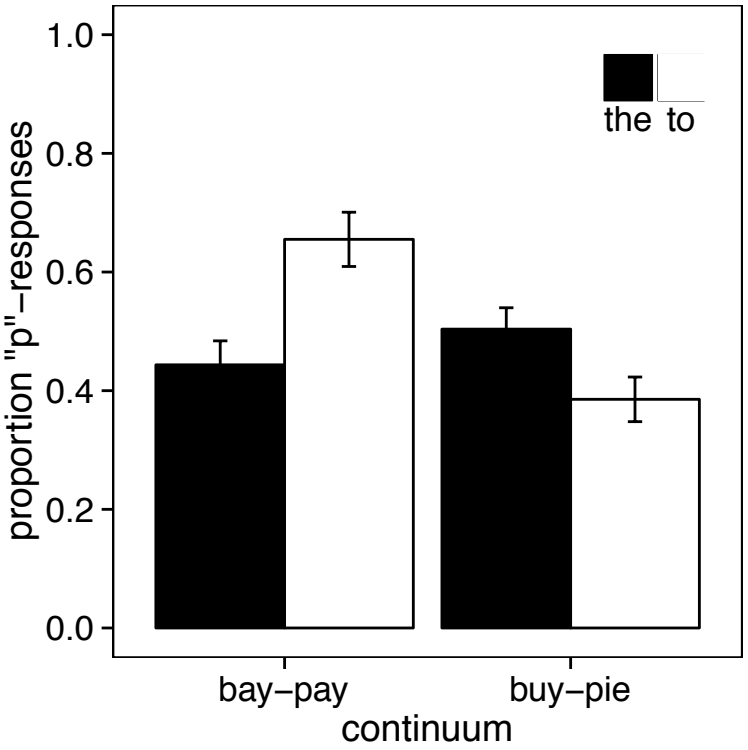

Figure 3.11. Reproduction of Figure 1.2. Mean proportion of /*p*/-responses to ambiguous tokens from each VOT continuum in Experiment 1 of Chapter 1 after noun-biasing and verb-biasing sentence contexts. Error bars represent standard error. (Fox & Blumstein, in press)

One possible explanation for this asymmetry lies in the strength of the biasing information in the prior of BIASES. Intuitively, if the targets (i.e., *bay*, *pay*, *buy*, *pie*) all represent relatively "good" nouns (i.e., they are sensible and/or grammatically acceptable following *the*), but only *pay* and *buy* represent "good" verbs (i.e., they are acceptable following *to*), then the second asymmetry should be predicted: the verb-biasing contexts

should drive the interaction. The stronger overall magnitude of the bias observed in the *bay–pay* continuum might occur under many circumstances, including if *pay* was particularly likely to follow *to* and *bay* was particularly unlikely, thereby creating a bias much stronger in that condition than in the others. To determine whether this intuitive explanation could quantitatively capture the asymmetry observed for these particular contexts and target words, we implemented the prior of BIASES (bigram language model; see Chapter 2) for these stimuli.

Table 3.4 provides corpus counts of the number of tokens of each of the function word / target bigrams (e.g., *to pay*, *the buy*) appears in the Google Books corpus (Michel et al, 2010). As described in Chapter 2, a smoothing parameter[6] is added to every corpus count (Lidstone, 1920) to yield an estimate of the conditional prior for each word, given the preceding context.

	...bay	**...pay**	**...buy**	**...pie**
to...	91,314	17,383,444	7,423,403	6,709
the...	3,236,957	945,799	56,284	249,243

Table 3.4. Number of tokens of each bigram found to the 2009 Google Books corpus (Michel et al, 2010)

Furthermore, the likelihood model was improved so as to allow the rime (i.e., vowel + glide) of the target stimulus to influence the likelihood of BIASES, rather than just the VOT of the initial stop consonant of the stimulus (see Chapter 4 for more detail on the mathematical details of this improvement). Words that differed from a target

[6] Although any value of alpha will give the basic same pattern of results (i.e., the same ordering of effect sizes), different values will accentuate the disparities between the contexts and continua to greater or lesser extents. For the present simulations the value, $1x10^7$ was utilized to illustrate the similarity between the model predictions and the experimental results.

stimulus in its rime were assigned a likelihood (and therefore a posterior probability) of zero; that is, for every trial on which a subject heard /**?***ei*/, the only words competing for recognition were *bay* and *pay*.

The smoothed corpus estimates were incorporated into BIASES as the conditional prior. For the likelihood model, arbitrary mean and variance parameters were selected from among those used in Simulation Study 3.1 (any choice shows the same basic pattern). Figure 3.12 shows the posterior probability distributions for the two continua in each context (left panels) and reproduces Figure 1.1 (right panels) for comparison.

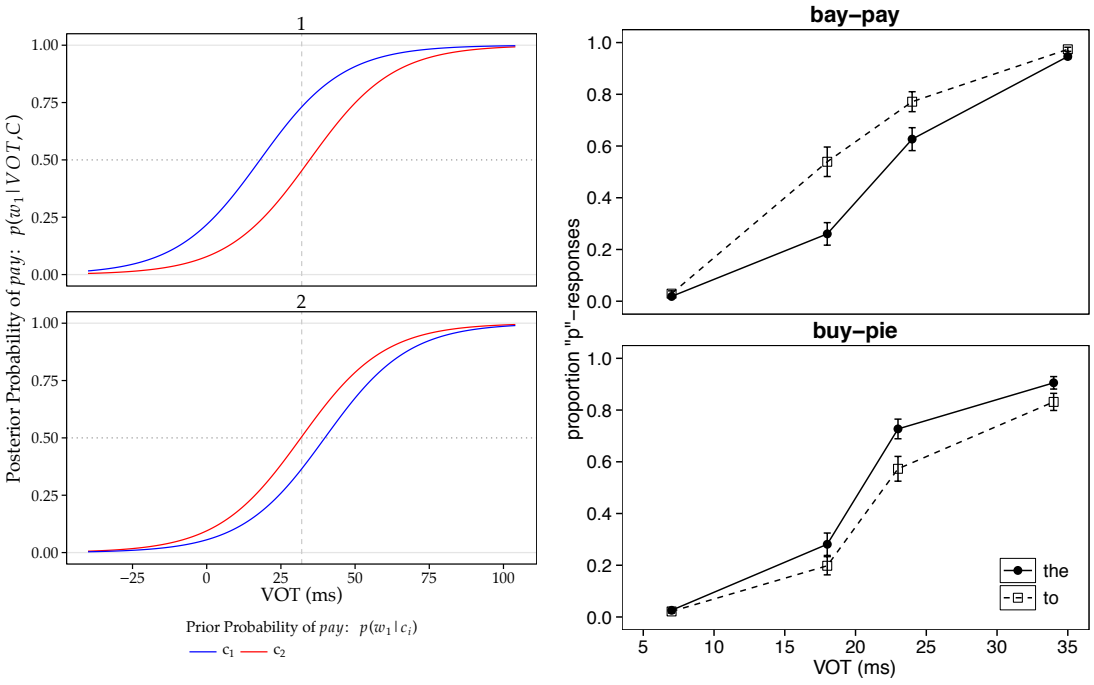

Figure 3.12. Model simulations of posterior probability distributions (left) and original data (right) for the *bay–pay* (top) and *buy–pie* (bottom) continua in the noun- and verb-biasing contexts (verb-biasing: blue on left / dashed on right; noun-biasing: red on left / solid on right). Right panels are a reproduction of Figure 1.1. Mean proportion of /*p*/-responses to tokens from each VOT continuum in Experiment 1 of Chapter 1 after noun-biasing and verb-biasing sentence contexts. Error bars represent standard error. (Fox & Blumstein, in press)

Finally, from the resulting posterior distributions, 20 sets (i.e., 20 subjects) of 20 behavioral responses were simulated in for each context condition for an ambiguous VOT (randomly selected VOT value within 5 ms of the assumed category boundary). Figure 3.13 shows the mean proportion of /*p*/-responses in each continuum to the ambiguous VOT value after each context. The same pattern of results seen in Figure 3.11 can be observed there: the overall interaction is robust, there is a larger effect of context in the *bay–pay* continuum than in the *buy–pie* continuum, and the effect is largely driven by the verb-biasing contexts. Importantly, these results are obtained without any significant efforts at parameter-fitting; rather, the pattern of results that emerges is inherent to a Bayesian model that assumes, like BIASES, that the prior word should bias the perception of subsequent spoken words when they are phonetically ambiguous. Thus, these results strongly suggest that variability in the strength of prior information in a sentence context modulates the size of observed top-down effects in systematic and predictable ways.

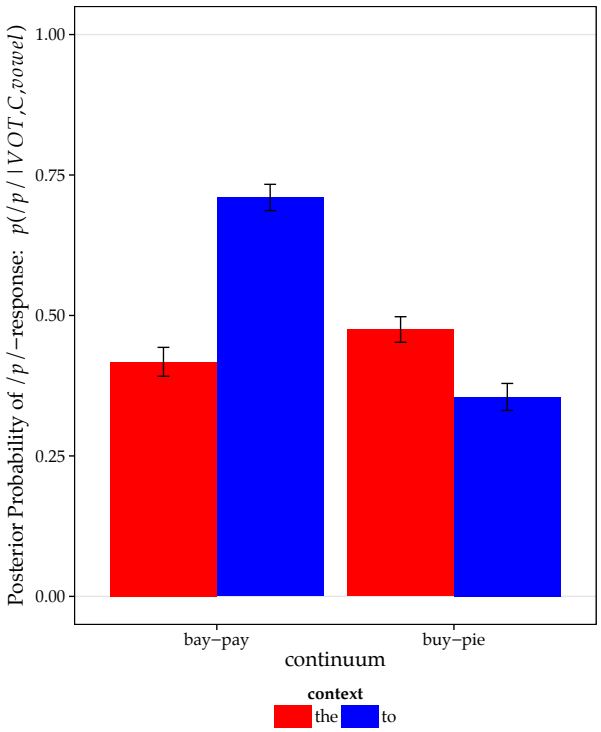

Figure 3.13. Model simulations of behavioral response rates for the *bay–pay* (left bars) and *buy–pie* (right bars) continua in noun-biasing (red) and verb-biasing (blue) sentence contexts. Mean proportion of simulated /*p*/-responses to randomly selected ambiguous VOT (within 5 ms of simulated category boundary) by 20 subjects with 20 Bernoulli (independent and identically distributed) trials. Error bars represent standard error of 20 simulated subject means. Compare to Figure 3.11 (or Figure 1.2). (*cf*. Fox & Blumstein, in press)

3.2.5. Variability in the Effect Sizes Compared to "Neutral" Prior Contexts

Another inconsistency that has received relatively little attention in the literature despite its appearance in various studies (e.g., Guediche et al, 2013; van Alphen & McQueen, 2001) relates to the size of top-down effects when responses to stimuli in biasing contexts are compared to a context that is designed to serve as a neutral condition. For instance, Guediche and colleagues (2013) examined responses to stimuli that were occasionally phonetically ambiguous between *goat* and *coat* after goat-biasing sentence contexts (e.g., *He milked the...*), coat-biasing sentence contexts (e.g., *He buttoned the...*), and neutral sentence contexts in which either goat of coat could sensibly serve as a

continuation (e.g., *He painted the...*). Researchers have also attempted to include similarly neutral contexts in lexical effect studies by using continua between two non-words or between two words (e.g., Fox, 1984). The goal of such studies is generally to compare each biasing context to the neutral context in order to illustrate that sentential context effects are affecting the identification of stimuli relative to a neutral context.

However, it has often been observed that the neutral context may differ significantly from only one of the biasing contexts, or the effect size may be larger in one direction than the other. These asymmetries have rarely been discussed in detail in the literature. Nonetheless, because BIASES demands that even the allegedly "neutral" context have some prior (whether 0.5 or not), this exercise serves as a reminder that one must explicitly model the prior on even the neutral context. It may be that the sentence contexts representing the neutral condition are, indeed, truly unbiased: $p(w_1|C = c_N) = 0.5$. In such a case, the asymmetries might be explained by asymmetric prior biases for the two biasing contexts: there is no guarantee that stimuli in the two biasing contexts will be equally biasing away from a perfectly neutral context, as in Simulation Study 3.1 (see Box 3.1) where $p(w_1|C = c_1) = 0.75$ and $p(w_1|C = c_2) = 0.25$. Indeed, in Simulation Study 3.2 (see Box 3.2), prior contexts that were not equally biased compared to the neutral context were examined, with asymmetries resulting (see Figures 3.6 and 3.9).

This issue is explored further later in this chapter, but, for the moment, it is simply worth noting that one important conclusion from this discussion is that an experimentally defined "neutral" context may not be neutral at all: there may be biases inherent in even those "neutral" stimuli, and – even if they are neutral – the only conditions under which

one should expect equal effect sizes of each context in comparison to the "neutral" condition is when stimuli are perfectly evenly biased around the perfectly neutral context. These conditions are unlikely to be met without extremely tight experimental design controls, but BIASES allows an experimenter to predict in advance of an experiment the likely effect sizes when comparing stimuli from different conditions. Thus, BIASES can be employed for power analyses and experimental design purposes.

3.3. Testing Predictions of BIASES: Experiment 3.1

In order to further examine the extent to which fine-grained predictions of BIASES could be observed in empirical data, a new set of stimuli was constructed and Experiment 3.1 was conducted. In particular, there were two goals: (a) to determine whether by-subject differences in *pay*-response rates to different acoustic stimuli predicted specific patterns of top-down effects, and (b) to determine whether there is evidence that subjects' responses to stimuli following a "neutral" context actually reflect Bayes-optimal processing. These two goals were addressed by two model comparison analyses examining the results of Experiment 3.1.

3.3.1. Methods

3.3.1.1. Subjects

15 healthy young adults participated in Experiment 3.1 as part of a multi-experiment session, although all 15 subjects completed this experiment first. Participants either received course credit or 8 dollars. All subjects were right-handed monolingual native speakers of American English, and all participants self-reported having normal hearing and no known neurological diseases.

3.3.1.2. Materials

The stimuli for this study were comprised of 4 acoustic tokens from a voice-onset time continuum between *bay* and *pay*, each of which was appended to a set of noun- and verb-biasing sentence contexts (e.g., *He hated the...* vs. *He hated to...*). Stimuli were recorded in a soundproof booth on an Edirol digital recorder (model R09-HR) with a Sony microphone (model ECM-MS907) (sampling rate: 44.1 kHz; 24 bits; stereo) and then were resampled in BLISS speech-editing software (Mertus, 1989) (sampling rate: 22.05 kHz; 16 bits; mono: left). The speaker was a male native speaker of American English. All sentence frames (e.g., *He hated...*), biasing function words (*to*/*the*), and naturally produced target tokens of *bay* and *pay* were produced in isolation multiple times and tokens were selected from among them for use in the experiment proper. The list of sentence frames consisted of the same 20 main verbs that were used by Fox and Blumstein (2015; see also Chapter 1), but first names were replaced with the pronoun "*He*" to reduce differences in stimulus duration and ensure subject would not be able to learn mappings between names and subsequent function words.

Three contexts were appended to each of the 20 sentence frames. A naturally produced token of *the*, a naturally produced token of t*o* (both of duration 125 ms), and 125 ms of unintelligible but spectrally similar speech babble (the initial and final 40 ms of which were ramp up/down respectively). This third condition was dubbed the "noise" condition. In total, this yielded 60 sentence contexts (20 main verbs crossed with the three conditions; *He hated to.../the...*/[*noise*]...). To each of these 60 contexts, each of 4 acoustically manipulated tokens from a VOT continuum between *bay* and *pay* were appended (yielding 240 total sentences for each subject to respond to in the experiment).

Tokens of the VOT continuum were constructed by concatenating: the unaltered burst of a *pay* token; a variable amount of aspiration from the natural *pay* token (duration depended on duration of vowel removed – see below); the first quasiperiodic pitch period from the natural *pay* token; and all but the first *N* pitch periods of a naturally produced token of *bay*, where *N* = the stimulus number - 1. The duration of the N pitch periods that were removed was equal to the amount of aspiration added from the *pay* token in order to ensure all tokens were the same duration, overall (within 1 ms of 439 ms). In this way, 7 VOT tokens were created. Four tokens with VOTs of 3, 22, 31, 48 ms were selected for stimuli because the middle two were judged to be the most ambiguous and other two were strong endpoint tokens.

3.3.1.3. Procedure

All subjects heard all stimuli binaurally over headphones in a random order in a sound-dampened booth and were instructed to respond whether the last word of each sentence was *bay* or *pay*, by pressing the appropriately marked button as quickly and accurately as possible, and to guess if they did not know. The buttons were counterbalanced across subjects. Subjects were told ahead of time that some sentences would not make sense. Subjects completed 2 practice trials before the experiment began. The experiment took about 12 minutes to complete. There were no breaks included during the experiment.

3.3.2. Results: Logistic Regression Analysis of Biased Contexts

The results for Experiment 3.1 are analyzed throughout the remaining sections of Chapter 3. First, we consider only the results for the two biasing contexts (shown in

Figure 3.14). Subsequent analyses examined the results of all three conditions (including noise).

To test for an effect of sentential context on speech recognition, the data were analyzed using mixed effects logistic regression (Baayen, Davidson & Bates, 2008; Jaeger, 2008) (see Chapter 1). There was evidence of a strong influence of VOT on the rate of *pay*-responses ($\beta = 0.31$, $p < 0.001$) and also a strong sentential context effect ($\beta = 3.03$, $p < 0.001$). Figure 3.15 shows the results by illustrating the effect size as a function of VOT.

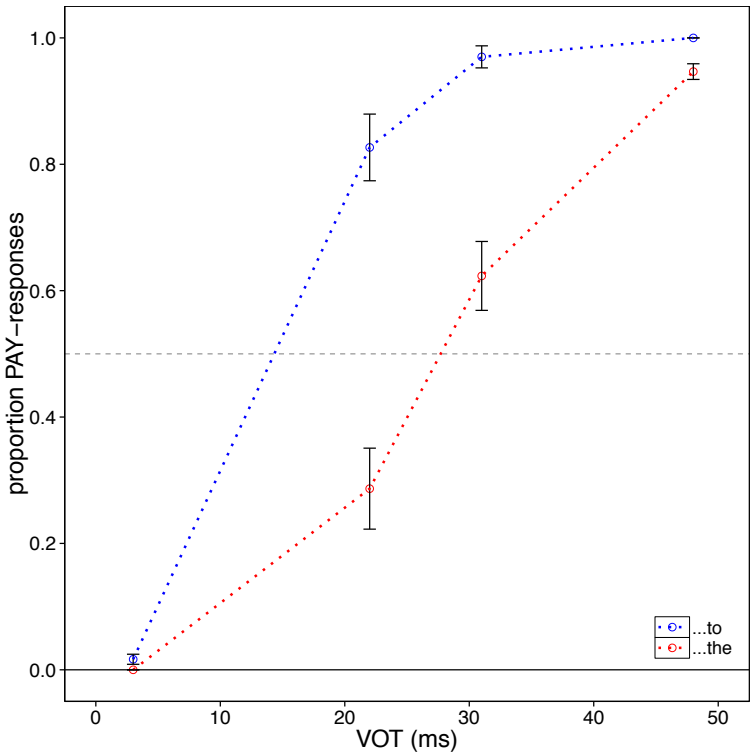

Figure 3.14. Mean proportion of *pay*-responses to tokens from the *bay–pay* VOT continuum after noun-biasing and verb-biasing sentence contexts. Error bars represent standard error.

Figure 3.16 shows the by-subject variability in effect sizes for responses to each VOT token. However, Figure 3.17 shows that subjects also vary in their underlying likelihood model. In particular, in their responses to these tokens, subjects' expected

category boundaries differ: some subjects appear to expect exemplars of *pay* to have much longer VOTs than other subjects. Because of this, we examined two models for BIASES: one in which subjects, all share the same category boundary, and one in which subjects differ. If subjects do, indeed, differ in their category boundary, then they should also vary in their expected effect size for a given VOT token.

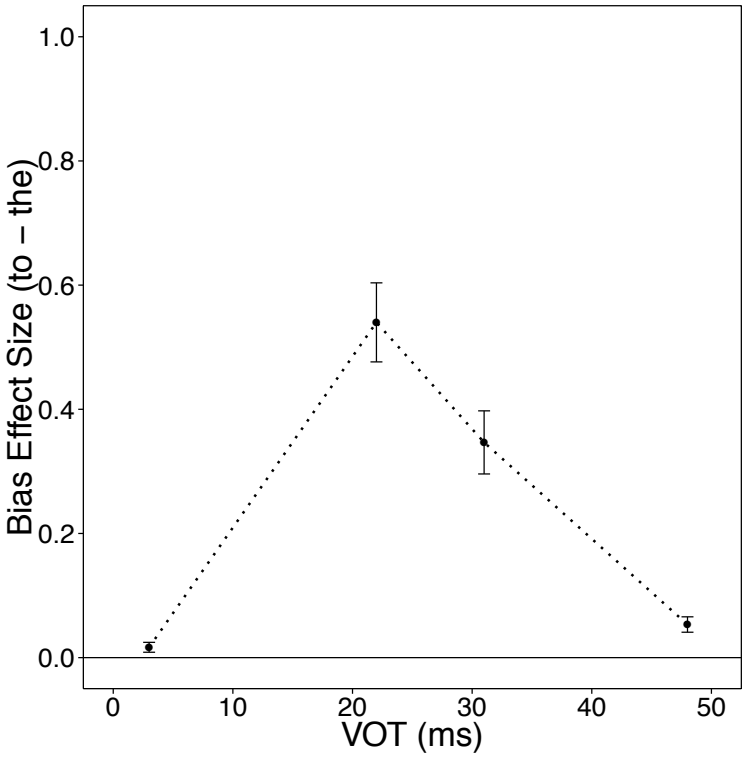

Figure 3.15. Mean difference in proportion of *pay*-responses to tokens from the *bay–pay* continuum after verb- vs. noun-biasing sentence contexts. Error bars represent standard error.

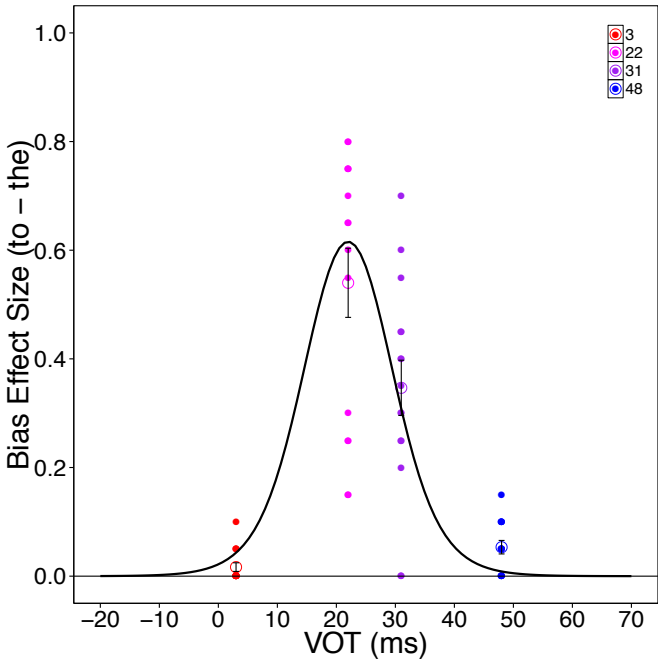

Figure 3.16. For each subject (N=15), difference in proportion of *pay*-responses to tokens from *bay–pay* continuum after verb- vs. noun-biasing contexts. Error bars represent standard error.

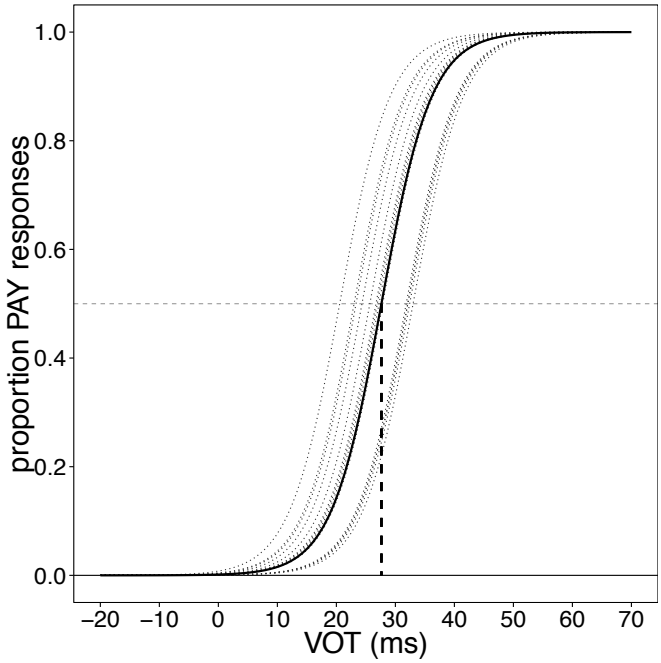

Figure 3.17. For each subject (N=15), their best-fitting (see section 3.3.2) unbiased posterior probability distributions (probability of *pay*-response to tokens from the *bay–pay* VOT continuum after a theoretical context that is truly neutral).

3.3.3. Results: Model Comparison 1 – Subject Variability

BIASES was implemented as a hierarchical Bayesian statistical model for further analysis. For the present analyses, only the data from responses to the VOT tokens after noun- and verb-biasing contexts (not in the noise context) were considered. Two separate versions of the model were implemented: in one version, subjects shared one group parameter for the mean of the normal likelihood function for their /*b*/ category and in the other version, the mean of the likelihood function could differ between subjects (the hyperprior on subjects' μ_B was presumed to be normally distributed).

As noted in Simulation Study 3.1 (see Box 3.1), because the current simulations of BIASES assume equal category variance (as do Feldman et al, 2009; Clayards et al, 2008; Kleinschmidt & Jeager, 2015), category variance and distance between category means are confounded. Thus, all model-fitting analyses assume a distance between categories based on VOT distributions from production data reported by Lisker and Abramson (1964): $\mu_P - \mu_B = 55\ ms$.

Tables 3.5 and 3.6 show the results of the model-fitting with and without by-subject variability, respectively. All chains converged, as judged from visual inspection of the chains and the Gelman-Rubin statistics for each model: multivariate psrf = 1.01 for both models and point estimates were all between 1.00 and 1.01 (with upper 95% confidence intervals of 1.00-1.03).

Critically, the DIC (popt) was computed for each model in order to determine whether the additional parameters allowing subjects to differ in their phonetic category structure improved the model fit significantly. Penalized deviance scores were 514 for the group-level model and 398.3 for the hierarchical model, despite having penalty terms of

6.602 and 44.99, respectively. Table 3.6 provides estimates and HDIs for the parameters in the hierarchical version of BIASES.

	Median	Mean	SD	95% HDI min	95% HDI max
α	0.089	0.090	0.016	0.059	0.120
σ^2	266.93	267.51	11.61	246.05	290.97
μ_B	0.15	0.44	0.43	-0.37	1.26

Table 3.5. Summary of posterior Markov chains from model that assumed group-level category structure (i.e., same μ_B for all subjects).

	Median	Mean	SD	95% HDI min	95% HDI max
α	0.060	0.061	0.012	0.040	0.085
σ^2	235.98	236.44	10.77	216.10	257.90
μ_{μ_B}	0.41	0.44	1.30	-2.12	3.04
$\frac{1}{\tau_{\mu_B}} = {\sigma_{\mu_B}}^2$	21.08	24.13	12.27	7.36	48.97

Table 3.6. Summary of posterior Markov chains from model that assumed hierarchical phonetic category structure (i.e., variable μ_B for subjects).

A posterior distribution was also obtained for each subject's μ_B, so we determined the median of each of these 15 posterior distributions and added half of the assumed $\mu_P - \mu_B$ (i.e., 27.5 ms) to compute a single point estimate for an approximate category boundary for each subject. Subjects' boundaries, calculated in this way, ranged between 20.72 ms and 33.37 ms (mean = 27.95, SD = 4.02). In order to illustrate the improved model fit obtained by hierarchical modeling of phonetic category structure, we computed the distance of each VOT token from the estimated boundary for each subject (calculated from the median of the subject's posterior distribution) and re-plotted Figure 3.16 with an x-axis reflecting the adjustment of subjects' boundaries to coincide at a single point. This can be seen in Figure 3.18. In short, subjects show larger effects when the model predicts that they should show larger effects (e.g., closer to the category boundary), and this fine-grained variability among subjects is neatly captured by BIASES' assumption that

subjects are not all identical in their phonetic expectations for acoustic realizations of exemplars of /*b*/ and /*p*/.

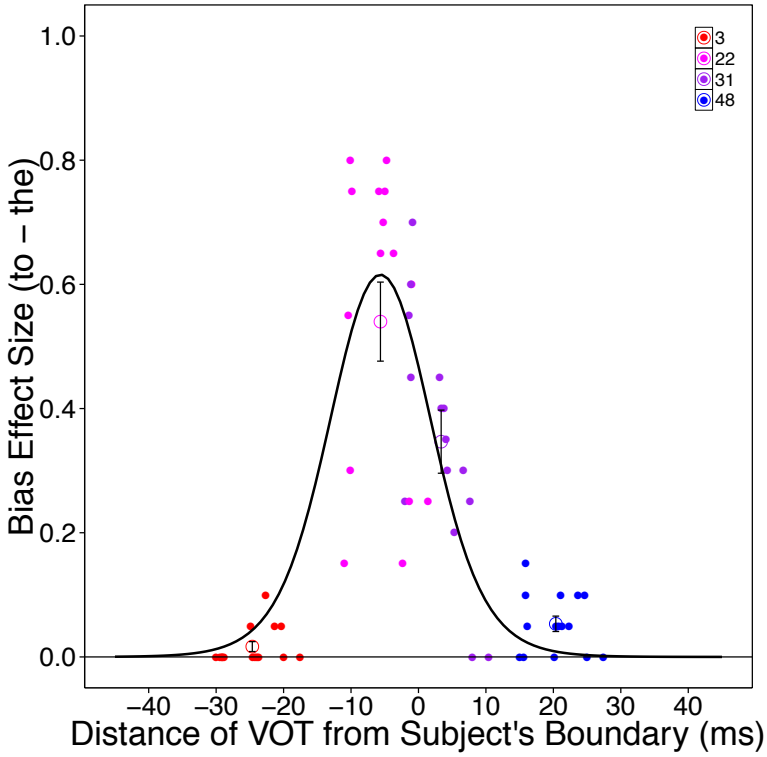

Figure 3.18. For each subject (N=15), the difference in the proportion of *pay*-responses to tokens from the *bay–pay* VOT continuum after verb-biasing vs. noun-biasing sentence contexts. Note that, unlike Figure 3.16 and others, the x-axis is adjusted for subjects' VOT boundaries. Error bars represent standard error.

3.3.4. Results: Model Comparison 2 – Inherent Biases in "Neutral" Priors

The previous analyses of the results of Experiment 3.1 have focused on the data from the noun-biasing and verb-biasing condition, but ignored the third "noise" condition in which subjects heard sentences like *He hated* [*noise*] */**?**ay/*.

As discussed earlier, when experimenters include a "neutral" condition, how subjects respond to stimuli in that baseline condition must be modeled just like subjects' responses to biased contexts. Next, we examined eight possible models of subjects' conditional priors in order to understand the principles underlying subjects' responses to

stimuli both in the biasing contexts and the noise condition employed in the current experiment.

If, in the noise condition, subjects are equally biased towards *bay* and *pay*, then the noise context would lie closest to the noun-biased sentence contexts because those noun-biased contexts are less biased, overall, than the verb-biased contexts (see Table 3.4). Note that, as discussion above, it is not likely that responses to the noise condition will lay perfectly midway between the noun- and verb-biasing contexts.

On the other hand, BIASES makes a different prediction about how subjects should respond in the noise condition. In particular, one principle of Bayesian models is that, when some information is not available, the optimal way to integrate that (lack of) information is to "believe" (in the Bayesian sense) each possible value of the cue to the extent that that cue was likely. This is called *marginalization* (see Chapter 2). In the present circumstances, this would mean that subjects' responses to stimuli in the noise condition should be closer to the verb-biasing contexts rather than the noun-biasing contexts. Thus, the "neutral" assumption and the Bayesian (marginalization) assumption make opposite predictions about where subjects' responses to stimuli in the noise condition should fall.

Figure 3.19 displays the results of subjects' responses to the noise condition added to the same data presented in Figure 3.14. As can be seen, responses in the noise condition lie closer to the verb-biased context than to the noun-biased context, suggesting that subjects were performing marginalization. However, to confirm this, we conducted a model comparison to evaluate the extent to which the marginalization model improved in fit over other alternative models.

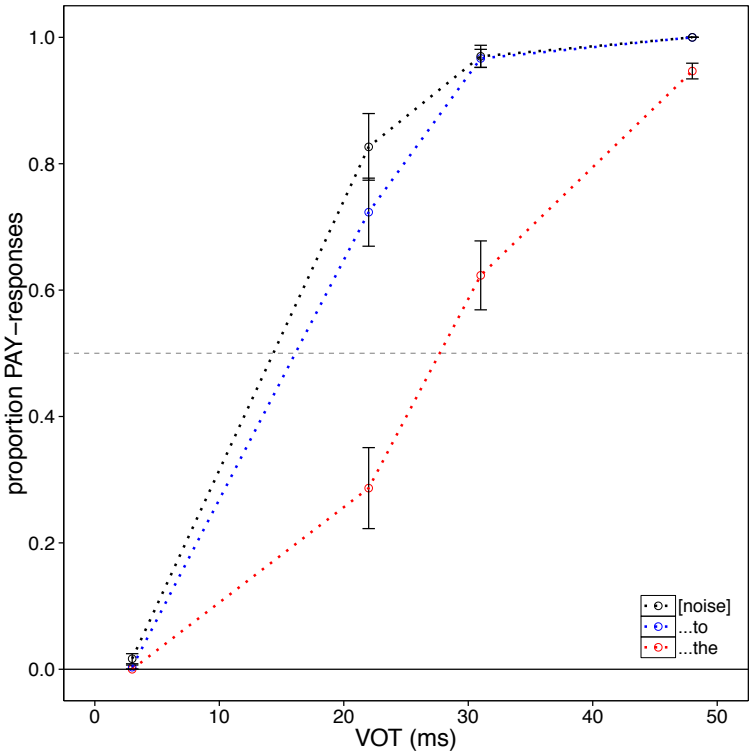

Figure 3.19. Mean proportion of *pay*-responses to tokens from the *bay–pay* VOT continuum following the noun-biasing (e.g., *He hated the...*) and verb-biasing (e.g., *He hated to...*) sentence contexts (see also Figure 3.14), as well as in the noise condition (e.g., *He hated* [*noise*]...). Error bars represent standard error.

Eight models were compared (see Table 3.7); they differed on the assumed prior for the context conditions (*to.../the...*) and the assumed prior for the noise condition. For four models, the priors for the context conditions were estimated based on the standard bigram model (described in Chapter 2), considering only probability of *bay* vs. *pay* based only on the preceding word (*to* vs. *the*). The remaining four models employed a more complex trigram language model for their priors, considering the probability of *bay* vs. *pay* based not only on the previous word (*to* vs. *the*) but also on the main verb preceding that. The four models of each type each employed a different model for the prior when subjects heard the target after the noise condition. These models varied in complexity. One model assumed that subjects treated *bay* and *pay* as equally likely after hearing the noise condition (Equal Priors). A second presumed subjects were biased based on the

lexical frequency of *bay* and *pay*. A third assumed that subjects considered the probability of *bay* and *pay* after marginalizing over the possible bigram contexts in the experiment (*to* and *the*). Finally, a fourth model assumed that subjects marginalized over all possible trigram contexts in the experiment (*to* and *the*, but only for after the main verbs in the study).

If subjects were performing optimally and making use of all information available to them, BIASES predicts that they should respond after the biasing contexts (*to.../the...*) based on the trigram prior, and that they should marginalize over all trigrams in Experiment 3.1. Indeed, the model comparison's results support that finding (see Table 3.7). Table 3.8 reports the best-fitting model parameters for this best-performing model.

Assumed Prior for Context Conditions	**Assumed Prior for Noise Condition**	**Mean Deviance**	**Penalty Term**	**Penalized Deviance**
Bigram (to.../the...)	Equal Priors	680.4	36.35	716.8
Bigram (to.../the...)	Lexical Frequency	517.3	41.87	559.2
Bigram (to.../the...)	Context-Sensitive (Bigram)	501.9	39.77	541.6
Bigram (to.../the...)	Context-Sensitive (Trigram)	699.9	41.14	741
Trigram (marginal)	Equal Priors	778.9	39.55	818.4
Trigram (marginal)	Lexical Frequency	578.1	40.75	618.9
Trigram (marginal)	Context-Sensitive (Bigram)	505.4	39.31	544.7
Trigram (marginal)	Context-Sensitive (Trigram)	474.5	40.31	514.8

Table 3.7. Summary of Model Comparison 2. Shaded row is best-fitting model, which was the model that assumed subjects make use of not only bigram contexts (i.e., the prior word), but also the second-back word (i.e., trigram contexts) in their contextual priors. This was the most detailed representation of contextual information tested.

	Median	Mean	SD	95% HDI min	95% HDI max
α	0.044	0.044	0.007	0.031	0.059
σ^2	214.71	214.83	7.89	199.96	230.83
μ_B	0.35	0.35	1.24	-2.10	2.74
$\frac{1}{\tau_{\mu_B}} = {\sigma_{\mu_B}}^2$	18.35	20.77	7.89	7.07	40.68

Table 3.8. Summary of posterior Markov chains from best-fitting model in Model Comparison 2 (shaded model in Table 3.7; fully trigram-driven context model). Note that all models assumed hierarchical phonetic category structure (i.e., variable μ_B for subjects).

3.3.5. Conclusion

In conclusion, the results of Chapter 3 suggest that listeners' behavior exhibits many hallmarks of a Bayesian spoken word recognition system. Overall, the results lend support to the validity and utility of BIASES for explaining and predicting subjects' speech recognition behavior in experimental tasks. BIASES is capable of accounting for a wide range of variability that is usually ignored by other computational models, including variability among subjects, variability due to speech cues other than VOT, and variability due to prior contexts in an experiment. These findings, along with the theoretical contribution of BIASES – as a model of speech perception in sentential context – illustrate the novelty of the present work. An even more powerful demonstration of the utility of BIASES, though, would be to leverage it to inform theoretical debates in psychology or neuroscience. One key goal of computational modeling is to advance scientific theory by testing and comparing competing hypotheses. This is the aim of Chapter 4.

Chapter 4

Top-Down Effects on Spoken Word Recognition in Aphasia:

A Model-Based Assessment of Information Processing Impairments

4.1. Introduction

4.1.1. Brief Introduction

In order to understand spoken language, a listener must ultimately map a continuous acoustic waveform onto discrete lexical forms, which stand at the interface between sound and meaning. However, spoken word recognition is not only influenced by the so-called *bottom-up* speech signal; as the signal undergoes higher-level cognitive processing, other information sources exert *top-down* influence on speech perception (for review, see Samuel, 2011). For instance, perception is lexically biased: a phonetically ambiguous segment between /*b*/ and /*p*/ tends to be identified as /*b*/ when followed by *–ash* (because *bash* is an English word, but not **pash*), but as /*p*/ when followed by *–ast* (where *past* is a word, but not **bast*) (Ganong, 1980). Similarly, perception is contextually biased: a phonetically ambiguous stimulus between two words (e.g., *bay* and *pay*) tends to be recognized as *bay* after a noun-biasing sentence context (e.g., *Valerie hated the...*), but as *pay* after a verb-biasing sentence context (e.g., *Brett hated to...*) (Fox & Blumstein, in press). Although the mechanisms underlying the integration of bottom-up and top-down cues remain the subject of considerable debate (see, e.g., McClelland, Mirman & Holt, 2006; Norris, McQueen & Cutler, 2000), there is no question that both types of information influence speech recognition.

Top-down effects on speech perception are of particular interest because they reflect dynamics at the confluence of perceptual and cognitive processing, so their

emergence and the characteristics of their distribution can reveal key insights about many aspects of human language function (see Chapter 3). Given the theoretical significance of this class of phenomena, it is noteworthy that far less is known about the pattern of such top-down effects in patients with aphasia. Most such individuals experience at least some receptive language impairments (Boller, Kim & Mack, 1977; Goodglass, Gleason & Hyde, 1970), with deficits arising at many different levels of processing (Goodglass, 1993; Lesser, 1978). What grants the status of top-down effects in patients with aphasia special importance, however, is the substantial evidence that lexical processing – that is, processing at the level where sound contacts meaning – is particularly vulnerable in aphasic syndromes (for review, see Blumstein, 2007).

For example, a classic finding about lexical processing by neurologically healthy adults is that, upon hearing a prime word (e.g., *cat*), the processing of a subsequent target that is a semantic associate of the prime (e.g., *dog*) is automatically facilitated relative to processing when the prime was not related (e.g., *table*), as measured by the time required to accurately decide that *dog* is a word (Meyer & Schvaneveldt, 1971). Moreover, the extent to which listeners access *cat* (and, in turn, the extent to which processing of *dog* is facilitated) is modulated by the acoustic (or phonological) distance from *cat* of a "mispronounced" prime, as indicated by the monotonic ordering of lexical decision latencies for *dog* after four different prime conditions (from fastest to slowest): *cat* < **gat* < **wat* < *table* (Milberg, Blumstein & Dworetsky, 1988a). Although the implicit processing of semantic associates of perceived primes is typically spared in aphasia (i.e., *cat* primes *dog*; Milberg & Blumstein, 1981; Milberg, Blumstein & Shrier, 1982), patients fail to exhibit the characteristic graded sensitivity observed in healthy adults

(Milberg, Blumstein & Dworetsky, 1988b), a result that has been interpreted as evidence for lexical processing deficits in such patients.

Nonetheless, it remains unclear what mechanisms are responsible for the observed dysfunctions (for review, see Mirman, Yee, Blumstein & Magnuson, 2011). One theory argues that lexical processing deficits arise directly from disruptions to processing dynamics at the level of the lexical representations themselves (Blumstein & Milberg, 2000; Janse, 2006; McNellis & Blumstein, 2001). However, in order to definitively conclude that lexical information is, indeed, specifically implicated, it is important to rule out the possibility that what appear to be lexical processing impairments are actually just downstream consequences of impairments in the bottom-up processing of the speech signal. Unfortunately, since auditory word processing must inevitably require both bottom-up speech processing and accessing the lexical representation, it is easy to see why it has been difficult to rule out this alternative explanation.

However, top-down lexical and contextual effects on speech perception may offer a unique window through which to view this question. For instance, the lexical effect (Ganong, 1980) taps information stored within lexical representations because it reflects a comparison between two phonologically similar interpretations of a stimulus, only one of which corresponds to a lexical representation. The existence of one representation (*bash*) in the lexicon and the corresponding absence of another (**pash*) conspire to bias subjects' identification of speech stimuli toward words. As such, to the extent that patients or groups of patients differ in the size of their lexical effect from controls, these differences might be taken to suggest disruptions arising at the lexical level itself. On the other hand, listeners' identification of spoken words and sounds should, of course, also be affected by

bottom-up phonetic and phonological processing deficits. Therefore, the pattern of lexical or contextual effects on speech perception in a patient with a bottom-up processing deficit might also be expected to diverge from the performance of healthy control subjects.

The critical question, then, is "When it comes to top-down speech processing, are there unique predictions about the expected consequences of 'virtual lesions' at different levels of the spoken word recognition system?" It is not necessarily intuitive how – even in a healthy speech processing system – phonetic, phonological, lexical and sentential processing levels interact during online speech perception and ultimately drive subjects' responses to, for instance, a stimulus that is ambiguous between *bash* and **pash*. This challenge is multiplied when attempting to deduce how disruptions at different processing levels or to specific cognitive mechanisms might affect the behavior of patients with brain damage and a complex constellation of symptoms at any (or potentially many) of those processing levels. Thus, it is difficult to generate clear predictions about expected patterns of top-down effects in patients with aphasia, and it is also difficult to draw any strong conclusions about the relationship between such data and the nature of those patients' fundamental information processing deficits, without first identifying a theoretical lens through which to view the data.

To that end, the present work enlists the *BIASES* model (*Bayesian Integration of Acoustic and Sentential Evidence in Speech*; Chapters 2-3), a probabilistic computational model of spoken word recognition that has been shown to successfully capture key aspects of top-down effects on speech perception in healthy adults. As we will show, BIASES makes clear predictions about how fine-grained differences in the size and

distribution of top-down influences from lexical and contextual cues should be expected to emerge as a function of which information-processing levels are disrupted. Thus, by examining the specific patterns of top-down effects from lexical and sentential context in patients with Broca's aphasia (BA) and Wernicke's or Conduction aphasia (W/CA), and comparing those results to the distribution of top-down effects in healthy controls, it is possible to distinguish between the independent contributions of a range of processing impairments (including at acoustic-phonetic, phonological, lexical, and contextual processing levels), even when multiple such impairments may coexist in a single patient or group of patients.

4.1.2. Overview of Chapter 4

The central aim of this chapter is to investigate the nature of top-down processing in patients with aphasia and to evaluate the extent to which the pattern of deficits observed in two groups – patients with Broca's aphasia (BA) and patients with Wernicke's or Conduction aphasia (W/CA) – might inform the broader theoretical question regarding the locus of patients' apparent lexical processing deficits. To that end, we further elaborate a Bayesian model of speech perception presented in earlier chapters, the *BIASES* model (*Bayesian Integration of Acoustic and Sentential Evidence in Speech*; Chapters 2-3). The fundamental principles embodied by this iteration of BIASES, which I call *BIASES-A* are consistent with its parent model, as discussed earlier. For example, preceding words can still bias a listener's identification of a stimulus via a context-dependent conditional prior, and the model's likelihood function still computes the relative fit of candidate representations given some acoustic values.

However, both the prior and likelihood terms of BIASES-A take different forms than they did when the model was introduced. These adaptations are critical for the model to address the question of theoretical interest here. In fact, in some ways, BIASES-A relaxes some of the assumptions present in the minimalist version of BIASES presented in Chapters 2-3. For instance, the drastically oversimplified likelihood function in BIASES, $p(A|w_i)$, characterized each word as a distribution over VOTs, implicitly ignoring subsequent cues (such as the rest of the word). This assumption was sufficient for modeling the perception of minimally paired words (e.g., *bay* vs. *pay*) that only differ as a function of VOT, but it must be updated in order to account for lexical biases arising as a function of subsequent phonological information (e.g., whether the rime of the word is *–ast* or *–ash*). Of course, while adding complexity to the model in this was improves its ability to accurately characterize the human speech processing system, relaxing certain assumptions requires committing to certain additional assumptions. However, most importantly, this approach illustrates with one of the key strengths of BIASES: its flexibility. The architecture of BIASES and its fundamental properties do not change when the prior and likelihood functions are updated to more accurately capture additional findings about human cognition and perception. Thus, while the main goals of this chapter are to assess the prevalence of top-down effects on speech perception in patients with aphasia and to address the theoretical question about the locus of lexical processing deficits in aphasia, this work also serves as a demonstration of the broad range of questions that BIASES can be leveraged to study. Chapter 4 is organized into 4 parts.

First, we briefly review the evidence for lexical processing deficits in aphasia, with special focus on two clinical groups – patients with Broca's aphasia (BAs) and

patients with Wernicke's or Conduction aphasia (W/CAs). Patients belonging to each group exhibit a unique pattern of lexical processing deficits. These results motivated the proposal of a theory referred to here as the *Lexical Activation Hypothesis* (Blumstein & Milberg, 2000; Janse, 2006; McNellis & Blumstein, 2001), which posits that the observed impairments emerge due to disruptions at the level of patients' lexical representations. However, as mentioned earlier, it is unclear whether the impairments might be fully accounted for by bottom-up processing deficits, which are known to afflict most patients with aphasia. Although relatively little is known about top-down processing in patients with aphasia, and although it is not necessarily obvious how top-down processing might be implicated in or affected by lexical processing deficits, we propose that a model-based analysis of top-down effects on speech perception may offer unique insights about the nature of lexical processing deficits, more broadly.

Second, we review the basic principles embodied by the BIASES model of speech perception and show how this probabilistic model can be theoretically linked to the Lexical Activation Hypothesis, which is predicated on a connectionist/activation-based approach to cognitive modeling. We update several assumptions and components of BIASES, calling this iteration *BIASES-A*, for *Aphasia*, highlighting the model's viability for not only capturing the fine-grained statistics of language function (see Chapter 3), but also its ability to reveal novel insights about the sometimes subtle details of language dysfunction. Alternatively, the *A* could stand for *Activation*, highlighting another critical contribution of this chapter: linking BIASES to more traditional (that is, connectionist) theories, models and approaches to thinking about spoken word recognition. We review the mathematical form of BIASES-A, briefly addressing the most important implications

of the changes from BIASES. Finally, we outline the present work's model-based approach to the characterization of top-down processing deficits in patients with aphasia.

Third, we present Simulation Study 4.1 and Experiment 4.1. Simulation Study 4.1 examines how information processing deficits at different levels should be expected to emerge in behavioral responses during an experiment testing for a lexical effect in patients with aphasia (Simulation Study 4.1). Experiment 4.1, conducted over two decades ago (Blumstein, Burton, Baum, Waldstein & Katz, 1994), examined the lexical effect in patients with aphasia. The present study's model-based reanalysis of its original data offers new insights into the specific deficits responsible for the patterns reported in the original study.

Fourth, Simulation Study 4.2 and Experiment 4.2 follow the same approach as was taken in Simulation Study 4.1 and Experiment 4.1, but they examine the sentential context effects examined in previous chapters. We argue that Chapter 4's computational and behavioral results, together, lend support to the key ideas embodied by the Lexical Activation Hypothesis.

4.1.3. Lexical Processing in Aphasia

4.1.3.1. Lexical Processing Deficits

Lexical access and spoken word comprehension is often profoundly disrupted in aphasia. Recall the early illustration of this by Milberg, Blumstein and Dworetsky (1988a, 1988b), who employed a lexical decision paradigm wherein subjects heard a prime-target pair and were instructed to decide whether the target was a word (e.g., *dog*) or a non-word (e.g., **jand*). On those trials for which the target was a word (*dog*), the prime that immediately preceded it could come from one of four categories: it could be

an unrelated word (*table*), a semantically related word (*cat*), a "close" mispronunciation of the semantically related prime (**gat*), or a "distant" mispronunciation of the semantic associate (**wat*). Healthy controls tend to correctly identify the target stimulus, *dog,* as a word fastest when it was immediately preceded by the correctly pronounced, semantically related prime (*cat*), followed in order of speed by **gat*, **wat*, and *table*, suggesting that lexical access to a word (and therefore to its semantic associates) is graded based on the phonological similarity of the input to the word (Milberg et al, 1988a; see also, e.g., Connine, Blasko & Titone, 1993; Connine, Titone, Dellman & Blasko, 1997; McMurray, Tanenhaus & Aslin, 2002; Utman, 1997; Utman et al, 2001).

In contrast, BAs exhibit semantic priming effects when *dog* is preceded by the correctly pronounced prime, *cat*, but they fail to show priming in either of the mispronunciation conditions. On the other hand, patients with W/CA are equally primed by **gat* and **wat* as they are by *cat* (Milberg et al, 1988b). These results have been interpreted as evidence that lexical access is disrupted in both patient groups, but that the nature of this disruption is not the same for all patients (see also Janse, 2006; Utman et al, 2001; Yee, Blumstein & Sedivy, 2008; but see, e.g., Del Toro, 2000; Tyler, 1992). In BAs, it is more difficult for bottom-up information to activate a lexical representation: only a very good perceptual match for *cat* is able to access the lexical-semantic network that must be engaged in order to facilitate subsequent recognition of *dog*. In W/CAs, though, even a poor match between the bottom-up signal and the stable phonological form of a word is able to access that word's meaning. Clearly, deviation from typical lexical processing dynamics in either direction is likely to impair word comprehension in the real world, where speech is noisy and error-laden (Dell, 1988; Vitevitch, 1997, 2002)

and words very often belong to dense phonological neighborhoods (e.g., *cat* is similar to *hat*, *bat*, *pat*, *rat*; Luce & Pisoni, 1998). Thus, the locus of lexical processing impairments is of great interest.

4.1.3.2. The Lexical Activation Hypothesis

What is the source of these lexical processing deficits? One theory holds that each group's impairment can be traced to the resting activation level of lexical representations (Blumstein & Milberg, 2000; Janse, 2006; McNellis & Blumstein, 2001). According to this perspective, referred to here as the *Lexical Activation Hypothesis*, the extent to which **gat* primes *dog* depends not only on the degree of phonological match between a **gat* and *cat*, but also on the baseline activation of *cat*. Consider a model of semantic priming in which activation spreads (*cf*. Collins & Loftus, 1975) from *cat* to *dog* only after the activation level of *cat* exceeds some propagation threshold (*cf*. Rumelhart, 1989), and, thereafter, the amount of priming is related to the amount of supra-threshold activation (up to some maximum activation level). McNellis and Blumstein (2001) showed such a model captures the graded priming results in healthy controls (Milberg et al, 1988a), and alterations to the resting activation levels could explain the patterns in BA and W/CA (1988b). Lower resting activation levels rendered it impossible for poorly matching input to exceed *cat*'s propagation threshold, thereby preventing semantic priming by both close (**gat*) and distant (**wat*) mispronunciations (as in BA), while raising resting activation levels caused *cat*'s activation not only to exceed its propagation threshold, but also to quickly reach its maximum level, yielding ceiling-level facilitation of recognition of *dog* following *cat*, **gat*, and **wat* (as in W/CA).

4.1.3.3. Alternative Accounts of Lexical Processing Deficits

Critically, the Lexical Activation Hypothesis posits that the locus of the lexical processing deficit is inherent to the lexical representation: words' resting activation levels are responsible for the observed impairment. However, some alternative explanations implicate the bottom-up processing of the speech signal and the time course of lexical activation. For example, the same pattern as was observed in W/CAs would be expected if those patients had perfectly normal lexical-level representations, but they sometimes misperceived **gat* and **wat* as *cat*. Since auditory comprehension is very frequently impaired in patients with W/CA (Blumstein, Baker & Goodglass, 1977a; Eggert, 1977; Luria, 1976; Robson, Keidel, Lambon Ralph & Sage, 2012), this possibility raises an important issue. Indeed, even though phonetic and phonological processing deficits are not as universally associated with BAs, virtually all patients, including BAs, appear to have at least some difficulties (Baker, Blumstein & Goodglass, 1981; Basso et al, 1977; Blumstein et al, 1977a, 1977b, 1984; Carpenter & Rutherford, 1973; Jauhiainen & Nuutila, 1977; Leeper, Shewan & Booth, 1986; Metz-Lutz, 1992; Miceli et al, 1978, 1980; Sasanuma et al, 1976; Utman et al, 2001; Yeni-Komshian & Lafontaine, 1983). This is consistent with neuroimaging research pointing to the involvement of both anterior and posterior brain regions in phoneme perception (Belin, Zatorre, Hoge, Evans & Pike, 1999; Blumstein, Myers & Rissman, 2005; Burton, 2001; Burton, Small & Blumstein, 2000; Poeppel, 1996).

Notably, Milberg and colleagues (1988b) did try to rule out this explanation. In a post-experiment lexical decision task, patients in both groups were shown to correctly reject **gat* and **wat* as non-words while also correctly accepting *cat* as a word. In line with this finding, many other studies also suggest that, generally speaking, individual

subjects' lexical processing deficits cannot be fully predicted by their bottom-up pre-lexical processing deficits alone (Baker et al, 1981; Basso et al, 1977; Blumstein et al, 1977a, 1977b, 1984; Carpenter & Rutherford, 1973; Caplan et al, 1995; Caplan & Utman, 1994; Csepe et al, 2001; Gow & Caplan, 1996; Jauhiainen & Nuutila, 1977; Leeper, Shewan & Booth, 1986; Metz-Lutz, 1992; Miceli et al, 1978, 1980; Sasanuma et al, 1976; Yeni-Komshian & Lafontaine, 1983). Nevertheless, Robson and colleagues (2012) argue that the primary deficit in W/CA is at the level of the phonological code (*cf*. Luria, 1976), suggesting that these studies' inability to find a significant correlation between W/CAs' phonological processing deficits and their other comprehension difficulties is due to unduly heterogeneous clinical populations, poor task selection, and other factors.

Additionally, it has also been suggested that Milberg and colleagues' (1988b) inability to detect priming in BAs following the mispronunciation conditions might have resulted not from an inherent disruption to lexical representations, but rather from a disruption to the *dynamics* (i.e., time course) of bottom-up lexical activation (Prather, Zurif, Love & Brownell, 1997; Swinney, Zurif & Prather, 1989; Swinney, Prather & Love, 2000). However, recent results using eye-tracking methodologies (which achieve more fine-grained temporal resolution than the priming paradigms) have disputed the idea that the time course of lexical activation is delayed in BA (for review, see Mirman et al, 2011; Yee et al, 2008).

4.1.3.4. Top-Down Effects and Lexical Processing

It is apparent that at least part of the theoretical bottleneck that has made the debate between bottom-up and lexical-level accounts of patients' lexical processing deficits difficult to resolve arises from the inherent difficulty in teasing apart bottom-up

processing which accesses lexical information and lexical-level information during typical word recognition tasks. For any task that evaluates behavioral responses to auditory words, potential lexical-level disruptions and potential downstream effects of bottom-up processing disruptions are necessarily confounded. However, top-down effects, are somewhat unique. Top-down effects measure the extent to which higher-level information sources – including lexical-level information (like lexical status or frequency) and contextual information that influences lexical-level predictions (*cf*. Chapter 2) – bias the perception of spoken words or sounds. What does it mean to observe a top-down effect for a given acoustic stimulus? If the same word-initial segment that is phonetically ambiguous between /*b*/ and /*p*/ is labeled /*b*/ when followed by *–ash*, more often than when followed by *–ast*, then this means that, for the same bottom-up stimulus, lexical-level information is influencing subjects' ultimate speech recognition (Ganong, 1980).

The significance of this observation is that the sizes of lexical or contextual effects are scaled with the strength of bias provided by top-down cues. However, bottom-up processing will also influence the ultimate size of the top-down effects: if the bottom-up processing reveals that a stimulus is almost certainly an exemplar of some particular word or phonetic category, then the top-down cue will have little impact on the response rate and there will not be a large top-down effect observed (*cf*. Chapters 2-3). Put another way, the ultimate size of a lexical or contextual effect on the perception of a stimulus will be influenced by both the bottom-up and the top-down processing of the stimulus (which includes lexical-level information and contextual information). As such, disrupting either bottom-up or lexical-level processing is likely to lead to behavioral differences in the size

of top-down effects on speech perception. The challenge is to separate out the influences of each. This theoretical and computational challenge is addressed from an information-processing standpoint by the *BIASES* model (*Bayesian Integration of Acoustic and Sentential Evidence in Speech*; Chapters 2-3), and it is to this model that we now turn.

4.2. Applying BIASES to Spoken Word Recognition in Aphasia

4.2.1. Brief Overview of BIASES

The *BIASES* model of speech perception describes the mathematically optimal way of combining top-down information sources (such as lexical frequency or the contextual predictability of a word) and bottom-up information sources (such as acoustic cues in the stimulus). In Chapter 3, it was shown that BIASES provides a principled account for a number of fine-grained differences in the sizes of top-down effects on speech perception, explaining how properties of the model's *prior* (which corresponds to top-down information sources) and the model's *likelihood* (which corresponds to bottom-up information sources) should influence the ultimate size of the top-down effect for a given pair of conditions (e.g., two sentential contexts) and for a given acoustic signal (e.g., for a given voice-onset time, or VOT). Critically, though, the model's predictions about how large a top-down effect should be depend on the information contained within a model's prior and likelihood functions. Thus, if the underlying information contained within either the prior or the likelihood term of BIASES is disrupted, or if the information processing dynamics that govern the computations within the prior or the likelihood term of BIASES are disrupted, the expected size of top-down effects for a given pair of contexts and a given acoustic stimulus will also change.

Thus, in order to gain some insight into the nature of the information processing deficits in aphasia, we adapted the parent model, BIASES, to allow for the examination of how different "virtual lesions" to a child model, *BIASES-A* (for *Aphasia*), should influence the predicted sizes of top-down effects from lexical status, from lexical frequency and from sentence contexts.

4.2.2. From Activations to Probabilities: Lexical Activation Hypothesis

BIASES is a probabilistic computational model which, like Shortlist B (Norris & McQueen, 2008), but unlike connectionist models of spoken word recognition (e.g., TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994; Merge: Norris, McQueen & Cutler, 2000), does not rely on any notion of activation. Instead, the amount of support for a given candidate in some set of mutually exclusive alternatives (e.g., a word in the lexicon) is related to its probability, which is computed relative to the other candidates. While this approach has many advantages (see, e.g., Chater, Tenenbaum & Yuille, 2006; Norris, 2006; Norris & McQueen, 2008) it is important to consider the relationship between probabilistic models and activation-based models (for recent reviews, see McClelland, 2009, 2013; McClelland, Mirman, Bolger & Khaitan, 2014).

This is particularly crucial for the present modeling effort because, while BIASES does not rely on any notion of activation, the theoretical claim instantiated within the Lexical Activation Hypothesis about the underlying basis of lexical processing deficits in aphasia is couched within the language of words' baseline levels of *activation*: BAs have lower baseline levels of activation than healthy adults, while W/CAs have higher baseline levels of activation than healthy adults. This raises the following critical question: what sort of lesion to the lexical information in BIASES would mimic the effects of changes in

baseline activation levels described by the Lexical Activation Hypothesis? The answer to this question requires drawing three theoretical links between activation-based models and probabilistic models.

First, note that real-valued activation levels in a finite set of units in a connectionist model can be scaled (i.e., each divided by the sum of the activations of the entire set) in order to create a probability distribution, and, critically, this computation preserves the relevant ratios of all pairs of activation levels (Hinton & Sejnowski, 1983; Khaitan & McClelland, 2010; Luce, 1959; McClelland, 1991; McClelland, Mirman, Bolger & Khaitan, 2014; Movellan & McClelland, 2001; for a tutorial and review, see McClelland, 2013). Second, lexical frequency (which, by definition, characterizes a probability distribution over the lexicon) has a robust effect on spoken word recognition and speech perception (e.g., Connine, Mullennix, Shernoff & Yellen, 1990; Dahan, Magnuson & Tanenhaus, 2001; Howes, 1954; Luce, 1986; Marslen-Wilson, 1987; Pollack, Rubenstein & Decker, 1960; Savin, 1963; Taft & Hambly, 1986). Applying the converse relationship described in the first theoretical link (connecting activation levels to probabilities), suggests that words' baseline lexical activations should be scaled by their lexical frequencies (see, e.g., Dahan, Magnuson & Tanenhaus, 2001). Thirdly, the last step is to determine how to mimic the raising and lowering of the baseline lexical activation of words in an activation-based framework? The approach we take is to transform the probability of each word w_i in the lexicon of N_w words by applying the function **A** (for **Activation**), which is defined in Equation 4.1:

Equation 4.1

$$\mathbf{A}(p(w_i), \phi) = \frac{p(w_i)^{\phi}}{\sum_{j=1}^{N_w} p(w_i)^{\phi}}$$

The function **A** raises each w_i's probability to the same exponent (ϕ), and then rescales the distribution so that it sums to one (as is required for any probability distribution). Crucially, **A** has the following four properties:

Property 1. For $\phi = 1$: $\mathbf{A}(p(w_i), \phi) = p(w_i)$ for all w_i

Property 2. For $\phi > 1$: $\mathbf{A}(p(w_i), \phi) < p(w_i)$ for less probable (initially) w_i

$\mathbf{A}(p(w_i), \phi) > p(w_i)$ for more probable (initially) w_i

Property 3. For $\phi < 1$: $\mathbf{A}(p(w_i), \phi) > p(w_i)$ for less probable (initially) w_i

$\mathbf{A}(p(w_i), \phi) < p(w_i)$ for more probable (initially) w_i

Property 4. For $\phi = 0$: $\mathbf{A}(p(w_i), \phi) = \frac{1}{w_i}$ for all w_i

When $\phi > 1$, the distribution becomes more extreme, or peaked, with the most probable words becoming even more probable and the least probable words becoming even less likely, so a virtual lesion that increases ϕ will cause the "rich to get richer." Conversely, when $\phi < 1$, the distribution becomes more uniform, essentially "watering down" frequency effects in the initial, un-lesioned distribution. Smaller values of ϕ reduce frequency effects further and further until $\phi = 0$, at which point frequency effects are totally eliminated by functionally transforming the prior distribution over words into the uniform distribution: $A(p(W), \phi = 0) = Unif(1, N_w)$.

4.2.2.1. Preliminary Simulations: Lexical Activation Hypothesis

In order to establish the theoretical link between BIASES-A and the Lexical Activation Hypothesis, we must determine how ϕ relates to changes in lexical activation levels. That is, will increases in the baseline lexical activation levels (as hypothesized to underlie the lexical processing deficits in W/CAs) more closely match Property 2 (the "rich get richer" case) or Property 3 (the "watering down" case)?

In order to match the Lexical Activation Hypothesis to the computational approach embodied by the function **A**, we simulated a simple example with a "toy lexicon" of N_w = 20 words. For each word w_i, a frequency f_i between 10 and 100 was randomly determined ($f_i \sim Unif(10,100)$) to serve as that w_i's effective baseline activation value, yielding a lexicon with activation values represented by the vector F (containing the frequencies for all N_w words). Then, in two separate simulations with the same lexicon F, to mimic the predicted activation levels for patients with BA and W/CA (McNellis & Blumstein, 2001), we either subtracted or added the same value, η, to each word's activation level, yielding a new baseline activation vector $F' = F \pm \eta$.[7] Finally, for both the activation subtraction simulation ($F'_{BA} = F - \eta$) and the activation addition simulation ($F'_{W/CA} = F + \eta$), the "pre-lesion" activation vector, F, and the updated "post-lesion" baseline activation vector, F', were normalized to obtain pre-lesion and post-lesion probability distributions (Equations 4.2-4.4):

Equation 4.2

$$p(W) = \frac{F}{\sum_{j=1}^{N_w} f_j}$$

Equation 4.3

$$p_{BA}(W) = \frac{F'_{BA}}{\sum_{j=1}^{N_w} f_j} = \frac{F - \eta}{\sum_{j=1}^{N_w} f_j - \eta}$$

Equation 4.4

$$p_{W/CA}(W) = \frac{F'_{W/CA}}{\sum_{j=1}^{N_w} f'_j} = \frac{F + \eta}{\sum_{j=1}^{N_w} f_j + \eta}$$

[7] To prevent any of the words' activation levels from becoming negative, η was set to half of the least frequent word's activation value ($\eta = \frac{\min(F)}{2}$).

Figure 4.1 shows the effects of subtracting and adding to the baseline activation levels for the corresponding probability distributions. Decreasing words' baseline activation levels (the mechanism implicated in lexical processing deficits in BA, according to the Lexical Activation Hypothesis) enhances frequency effects. The most frequent words (i.e., words with relatively higher pre-lesion activation levels) became even more probable, and the rarest words became even less probable. Conversely, increasing words' baseline activation levels (the mechanism implicated in lexical processing deficits in W/CA, according to the Lexical Activation Hypothesis) diminishes frequency effects. The most frequent words and the least frequent words have less disparate post-lesion probabilities.

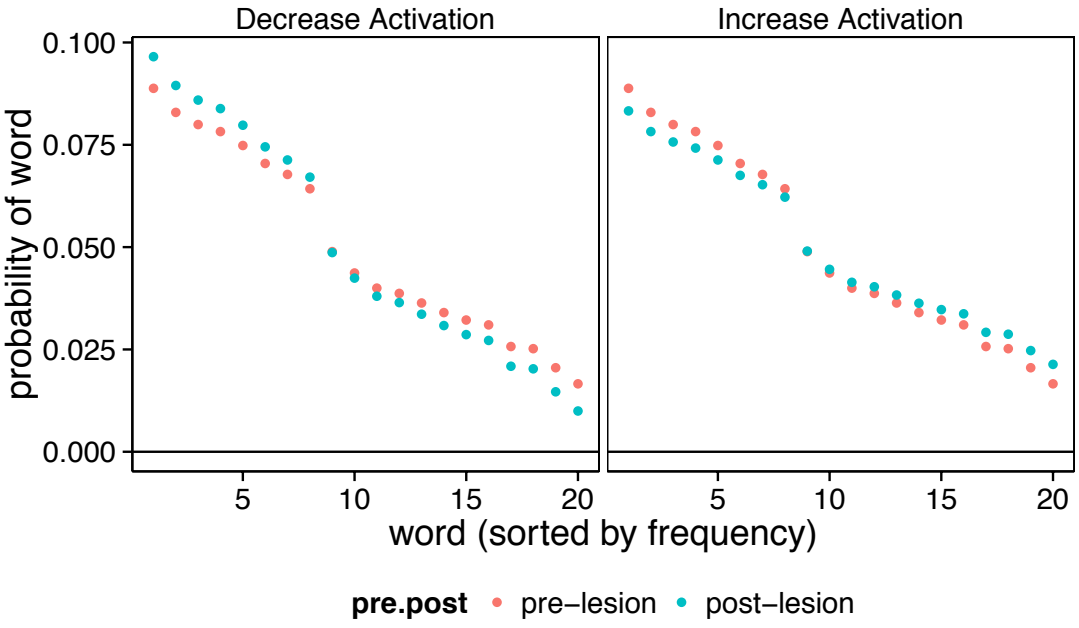

Figure 4.1. Results of simulations of Lexical Activation Hypothesis: the probability of each word before and after the virtual lesion. Virtual lesions involved either increasing or decreasing the activation level of each word by a constant amount (*cf.* McNellis & Blumstein, 2001).

Connecting the Lexical Activation Hypothesis and this simulation's results back to the probabilistic model (BIASES-A) and the function **A**, it is clear that decreasing the baseline activation levels of words corresponds to increasing ϕ (see Property 2 of **A**), while increasing baseline activation levels of words correspond with decreasing ϕ (see Property 3 of **A**). At an intuitive level, if baseline activation levels are increased by a constant amount (as in W/CAs; Blumstein & Milberg, 2000) while leaving the threshold for lexical access or word recognition constant (McNellis & Blumstein, 2001), and thereby requiring less bottom-up activation to achieve lexical access (i.e., **wat* primes *dog* as much as *cat* in W/CA; Milberg et al, 1988b), then the activation of the lexical representation will not be a very reliable cue to the actual presence of the word in the speech signal. Consequently, lexical-level cues should be less reliable for W/CAs than they are for healthy adults; from an information processing perspective, less reliable cues should be down-weighted, which is precisely the effect of decreasing ϕ. The opposite is true of decreasing baseline activation levels and increasing ϕ: the eventual activation of a lexical representation in BA is a more reliable cue to the actual presence of the word in the speech signal, leading to greater reliance on lexical-level information.

Note that, while our approach in the activation-based simulations above was designed to match the approach taken by McNellis and Blumstein (2001) in their proof-of-concept computational implementation of the Lexical Activation Hypothesis (adding/subtracting a constant value η to each word's activation level; *cf*. Morton, 1969; see also Norris, 2006), this approach is not mathematically equivalent to **A**'s exponential re-weighting of the entire distribution by ϕ. Our central aim was to match the overall directionality of the effects of parametric manipulations in the activation-based and

probabilistic frameworks on extreme values (e.g., the most and least frequent words). It is also worth noting that McNellis and Blumstein (2001) did not explicitly account for frequency effects. What is most important, though, is that while the details of the present probabilistic approach are not identical to the approach taken by McNellis and Blumstein (2001), the overall theoretical link is clear: the Lexical Activation Hypothesis should predict that, if controls have $\phi = 1$, then behavioral responses in BAs should be more influenced by lexical-level information ($\phi > 1$), while W/CAs should exhibit less of an influence from lexical-level information ($\phi < 1$).

4.2.2.2 Implications for Top-Down Effects on Speech Perception

Having derived the theoretical implications of the Lexical Activation Hypothesis for the probabilistic modeling approach, it is now possible to deduce principled predictions about the influence of lexical-processing deficits on top-down processing of speech. Because the Lexical Activation Hypothesis, as interpreted here, predicts that lexical-level information will be weighted *more* by BAs than by healthy controls, but *less* by W/CAs than by healthy controls, lexical status should have a greater influence on BAs' responses to stimuli ambiguous between a word and non-word, but it should have a weaker influence on W/CAs' responses. Implicit in this conclusion is the assumption of a relationship between "lexical status" and "lexical frequency." This assumption represents the basic principle that non-words are, in the limit, not so different from very low frequency words. Thus, the effects of lexical status and frequency effects are given a unified, if simple, explanation within the prior of the BIASES model: listeners expect to encounter more probable stimuli (see Chapters 2-3). Since non-words are less probable than words (see also, Norris & McQueen, 2008), an effect of lexical status might be

thought of as a special case of a lexical frequency effect. It is worth noting that since Ganong's (1980) original demonstration of the lexical effect, a number studies have reported hints of frequency effects (e.g., Fox, 1984; Fox & Blumstein, in press; Newman et al, 1997) and Connine, Blasko and Titone (1993) showed that the frequency of words within an experiment could drive top-down effects on speech perception that mirrored Ganong's lexical effect (1980).

Note, however, that, since lexical frequency is estimated by counting the number of times a word appears in some corpus, all non-words will, by definition, have a lexical frequency of 0. Clearly, subjects must be capable to recognize a string of phonemes that they have never heard before. While most speech an adult will hear on any given day will be composed of words in her lexicon, there must be some mechanism to "back off" to when a listener encounters foreign words, new words, or proper nouns such as names they have never heard before. Additionally some such computational machinery is obviously critical for learning in infancy and childhood (Feldman, Griffiths, Goldwater & Morgan, 2013). Even more relevant to the current situation, in the context of an experimental setting like Ganong's (1980), in which subjects hear dozens or sometimes hundreds of trials, they often identify the stimulus as a non-word. Thus, a subject's prior expectation should certainly not be completely determined by the lexical frequency of a stimulus as estimated from a corpus. In order to account for top-down effects of both lexical status and lexical frequency, what is needed is a prior that is influenced by frequency, but which also allows subjects to "expect the unexpected," as the case may be for non-words (with a corpus-estimated frequency of 0).

Thus, in order to account for the effect of lexical status on spoken word recognition within the probabilistic framework, a simple approach is to allow non-words have some small prior probability (Norris & McQueen, 2008; see also Chapters 2-3). In the current model, the prior probability for a non-word is estimated by fitting a smoothing parameter (Lidstone, 1920) to healthy controls' lexical effect data. Based on the success of parallel assumptions about BIASES in modeling the sizes of sentential context effects (see Chapter 3), we assume that control subjects are optimally making use of lexical information ($\phi = 1$), but that their lexical prior includes lexical frequency information that is smoothed by some positive and nonzero "pseudo-count" (α), which serves as an estimate of the prior expectancy for all non-words. It is further assumed that patients' underlying model of lexical information is the same as the controls' model (i.e., the same frequency counts for all words and the same smoothing parameter, α), but that patients may weight this information differently than the controls (ϕ).

It is important to note that drawing a relationship between lexical status and lexical frequency does not demand that every possible non-word be explicitly represented in the mental lexicon alongside every word, albeit with a different (lower) effective frequency estimate; indeed, this would not be a very plausible lexicon. Rather, we adopt the theoretical perspective that, during speech perception, candidate word-forms compete for recognition in a *lexical buffer* (Blumstein, 1994, 1998). A candidate's prior probability is determined by the sum of α (the smoothing parameter discussed above) and a candidate's lexical frequency (0 for non-words, but non-zero and positive for words). This framework allows all candidates (whether words or non-words) to have some baseline probability of being perceived (related to α), while a word's prior is also

influenced by its frequency. Because a constant α is added to the counts of all words, the prior probability of any word (with a nonzero frequency estimate) is greater than that of a non-word, allowing the model to capture top-down effects of lexical status (Ganong, 1980), and the prior probability of a given word is greater than that of any less frequent word, allowing the model to capture top-down effects of lexical frequency (Connine et al, 1993). Moreover, this framework predicts that the relative size of shifts in categorization due to lexical status should be more apparent the more common the word in a non-word/word pair (e.g., a greater bias towards *past* in **bast–past* than towards *bash* in **bash–pash*) is. We return to this prediction later.

The conclusions of this section are summarized in Table 4.1.

Clinical Diagnosis[a]	**Semantic Priming Lex. Dec. Latencies**[b]	**Baseline Lex. Act.**[c]	$p(W)$ **weight**[d]	**Lexical Effects**[e]	**Frequency Effects**[f]
W/CA	*cat*=**gat*=**wat*<*table*	$\rho_{W/CA} > \rho_C$	$\phi < 1$	$\lambda_{W/CA} < \lambda_C$	"watered down"
Control	*cat*<**gat*<**wat*<*table*	ρ_C	$\phi = 1$	λ_C	typical
BA	*cat*<**gat*=**wat*=*table*	$\rho_{BA} < \rho_C$	$\phi > 1$	$\lambda_{BA} > \lambda_C$	"rich get richer"

Table 4.1. Summary of Probabilistic Approach to the Lexical Activation Hypothesis: [a] W/CAs = Wernicke's or Conduction Aphasia; BA = Broca's Aphasia; [b] Semantic Priming Lexical Decision Task; patterns of response latencies for YES responses to dog (Milberg, Blumstein & Dworetzky, 1988); [c] Pattern of resting activation levels (ρ) according to Lexical Activation Hypothesis (Blumstein & Milberg, 2000; McNellis & Blumstein, 2001); [d] Pattern of weighting of lexical information (ϕ) in probabilistic approach that matches effects of baseline lexical activation modulation; see Equation 1 and Figure 1; [e] Predicted effects of model lesion on the influence of lexical status (λ); based on ϕ; [f] Predicted effects of model lesion on the influence of lexical frequency information; based on ϕ

4.2.3. Implementing BIASES-A

As described in Chapter 2, the fundamental assumption of BIASES is that when subjects categorize speech stimuli, their responses reflect both (1) the relative perceptual match between the available acoustic signal and each candidate response, and (2) the relative predictability of each candidate, irrespective of what was ultimately perceived. In

that sense, BIASES reflects the optimal integration of a bottom-up, perceptually driven processing stream and a top-down, expectation-driven processing stream. The introduction of BIASES in Chapters 2 and 3 places much focus on modeling the top-down constraints on expectations for future words provided by sentential context. However, as discussed above (and in Chapter 2), lexical status and lexical frequency can also serve to constrain expectations for upcoming linguistic material (*cf.* Norris & McQueen, 2008). Equation 4.5 encapsulates the fundamental properties of BIASES:

Equation 4.5

$$p(w_i|A) \propto p(A|w_i)p(w_i)$$

In short, subjects' word identification decisions are generated based on the posterior probability function, $p(W|A)$, which is proportional to the product of the likelihood, $p(A|W)$, and the prior, $p(W)$. While the likelihood function indexes how representative of each candidate word the perceived speech signal is, the prior indexes how probable each candidate word was to begin with (or, *a priori*). Also, recall that BIASES implements the influence of sentential context, *C*, by allowing *C* to constrain the prior, $p(W|C)$.

While the basic form of Equation 4.5 also underlies *BIASES-A* (the "child model" presented in this chapter), several adaptions were made to BIASES, the effects of which were (a) to enhance the breadth of the empirical coverage of BIASES, and, importantly for the questions addressed in this chapter, (b) to leverage the theoretical framework provided by BIASES for the purpose of providing a computational-level explanation for fine-grained differences in the patterns of top-down effects in patients with aphasia. In doing so, several simplifying assumptions made during the initial presentation of

BIASES were revisited, ultimately yielding a more complicated, but also more realistic and more accurate, model of human speech perception.

4.2.3.1. Adapting the Prior and Likelihood of BIASES

The model was enhanced in four main ways. First, the likelihood function was updated to include a phonological-processing stage interceding between the acoustic and lexical level. Second, the likelihood model was updated in another way to allow the rime of a stimulus to influence speech recognition, rather than only accounting for acoustic information available from the onset's VOT. Third, a smoothing parameter was added in order to allow novel phonological forms to have nonzero prior probabilities. Finally, a lexical buffer was added to the model, following work suggesting that the phoneme identification task may not necessarily tap phonemic processing, *per se* (Fox & Blumstein, in press; Swinney & Prather, 1980).

In the updated model, *BIASES-A*, upon perceiving a monosyllabic stimulus, Bayes' rule gives the probability of recognizing a candidate word-form, f_i, given the initial segment's voice-onset time, V, and the stimulus's rime, R (Equation 4.6).

Equation 4.6

$$p(f_i|V,R) = \frac{p(f_i)p(V,R|f_i)}{\sum_{j=1}^{N_f} p(f_j)p(V,R|f_j)}$$

Assuming that the VOT and rime are independent cues to the phonological form of the stimulus yields Equation 4.7:

Equation 4.7

$$p(f_i|V,R) = \frac{p(f_i)p(R|f_i,V)p(V|f_i)}{\sum_{j=1}^{N_f} p(f_j)p(R|f_j,V)p(V|f_j)} = \frac{p(f_i)p(R|f_i)p(V|f_i)}{\sum_{j=1}^{N_f} p(f_j)p(R|f_j)p(V|f_j)}$$

In the current model, and as shown in Equation 4.8, we assume rimes to be deterministically related to word-forms; while this is certainly not true of real speech, the lexical effect stimuli examined in this study (see Experiment 4.1's Methods) were blocked by continuum, so subjects only heard stimuli with a single rime many times for half of the experiment, and then the rime switched for all of the stimuli for the rest of the experiment. Thus, for our purposes, it is probably safe to assume that participants could accurately map rimes to associated word-forms. In particular, the lexical effect continua considered here ranged from **dut* to *toot* and *duke* to **tuk* (see Experiment 4.1's Methods).

Equation 4.8

$$p(r_{/ut/}|f_i) = \begin{cases} 1 & f_i \in \{/tut/,/dut/\} \\ 0 & f_i \in \{/tuk/,/duk/\} \end{cases}$$

$$p(r_{/uk/}|f_i) = \begin{cases} 0 & f_i \in \{/tut/,/dut/\} \\ 1 & f_i \in \{/tuk/,/duk/\} \end{cases}$$

Add-alpha smoothing (Lidstone, 1920) was implemented for the prior. Thus, even word-forms that never appeared in the Brown corpus (Kucera & Francis, 1963) had some prior probability. Equation 4.9 indicates the smoothed frequency estimates for all four relevant word forms (prior to normalization) in Experiment 4.1, where κ_{toot} and κ_{duke} are lexical frequency counts of those words in the Brown corpus and α is the smoothing parameter, fit as described earlier.

Equation 4.9

$$p(f_i) \propto \begin{cases} \kappa_{toot} + \alpha & f_i = /tut/ \\ \alpha & f_i = /dut/ \\ \alpha & f_i = /tuk/ \\ \kappa_{duke} + \alpha & f_i = /duk/ \end{cases}$$

The second term of the likelihood function, $p(V|f_i)$, which mapped VOTs onto word-forms, was modeled such that the mixture components of the Gaussian mixture model were onsets (namely, either /*t*/ or /*d*/) with normally distributed VOTs (still assuming equal category variance). As a simplifying assumption, it was assumed that the distribution of VOTs for a word's initial consonant depends on which consonant the word begins with (/*t*/ vs. /*d*/), but is otherwise independent of the identity of the word itself (but see, e.g., Baese-Berk & Goldrick, 2009; Fox, Reilly & Blumstein, 2015). Equations 4.10 – 4.12 show how $p(V|f_i)$ is expanded to account for a phonological processing level.

Equation 4.10

$$p(V|f_i) = \sum_{k=1}^{N_o} p(V, o_k|f_i) = \sum_{k=1}^{N_o} p(o_k|f_i)p(V|f_i, o_k) = \sum_{k=1}^{N_o} p(o_k|f_i)p(V|o_k)$$

Equation 4.11

$$p(V|f_i) = p\left(o_{/t/}\middle|f_i\right)p\left(V\middle|o_{/t/}\right) + p\left(o_{/d/}\middle|f_i\right)p\left(V\middle|o_{/d/}\right)$$

Equation 4.12

$$V|o_k \sim N(\mu_k, \sigma_k^2)$$

$$p(V|o_k) = \frac{1}{\sqrt[2]{2\pi\sigma_k^2}} e^{-\frac{(V-\mu_k)^2}{2\sigma_k^2}}$$

Additionally, a parameter was added that allows perceptual processing of the onset's VOT to be degraded, with the degree of degradation assumed to be independent of the value of the onset's VOT, as shown in Equation 4.13.

Equation 4.13

$$S|v \sim N(v, \sigma_N^2)$$

$$S|o_k \sim N(\mu_k, \sigma_k^2 + \sigma_N^2)$$

$$S|o_k \sim N(\mu_k, \sigma^2 + \sigma_N^2)$$

4.2.3.2. Modeling Speech Processing Deficits in BIASES-A

In addition to the assumptions outlined above, BIASES-A makes three basic assumptions, listed in Equation 4.14, about speech processing in healthy adults (young controls and age-matched [to the patients] controls). First, the relationship between word-forms and their onsets is assumed to be deterministic. Secondly, we assume that σ_N^2, the additional variance associated with perceptual processing deficits, is 0 for all healthy control subjects. That is because, when equal category variance is assumed, unless a noise manipulation is included in the experiment (*cf*. Feldman et al, 2009), σ_N^2 and σ_k^2 are not identifiable parameters in model-fitting. Thirdly, as mentioned earlier, we assume that healthy controls optimally weight lexical information after fitting a smoothing parameter, α, to the model.

Equation 4.14

$$\varepsilon_{YC} = \varepsilon_{AMC} = 0$$

$$\sigma_{N_{YC}}^2 = \sigma_{N_{AMC}}^2 = 0$$

$$\phi_{YC} = \phi_{AMC} = 1$$

Critically, these three assumptions were not made about speech processing in patients with aphasia. To the extent that patients do not have perfect lexical-phonological processing, BIASES-A implements this as shown in Equations 4.15 – 4.17.

Equation 4.15

$$p(o_{/t/}|f_i) = 1 - p(o_{/d/}|f_i) = \begin{cases} 1-\varepsilon & f_i \in \{/tut/, /tuk/\} \\ \varepsilon & f_i \in \{/dut/, /duk/\} \end{cases}$$

Equation 4.16

$$V|f_i \sim \begin{cases} (1-\varepsilon)\cdot N(\mu_{/t/}, \sigma^2_{/t/}) + \varepsilon \cdot N(\mu_{/d/}, \sigma^2_{/d/}) & f_i \in \{/tut/, /tuk/\} \\ \varepsilon \cdot N(\mu_{/t/}, \sigma^2_{/t/}) + (1-\varepsilon)\cdot N(\mu_{/d/}, \sigma^2_{/d/}) & f_i \in \{/dut/, /duk/\} \end{cases}$$

Equation 4.17

$$V|f_i \sim \begin{cases} \frac{1}{\sqrt[2]{2\pi\sigma^2}} \cdot \left[e^{-\frac{(V-\mu_{/t/})^2}{2\sigma^2}} + \varepsilon \cdot \left(e^{-\frac{(V-\mu_{/d/})^2}{2\sigma^2}} - e^{-\frac{(V-\mu_{/t/})^2}{2\sigma^2}} \right) \right] & f_i \in \{/tut/, /tuk/\} \\ \frac{1}{\sqrt[2]{2\pi\sigma^2}} \cdot \left[e^{-\frac{(V-\mu_{/d/})^2}{2\sigma^2}} + \varepsilon \cdot \left(e^{-\frac{(V-\mu_{/t/})^2}{2\sigma^2}} - e^{-\frac{(V-\mu_{/d/})^2}{2\sigma^2}} \right) \right] & f_i \in \{/dut/, /duk/\} \end{cases}$$

Figure 4.2 illustrates simulations teasing apart the influence of lexical-phonological processing impairments ($\varepsilon > 0$) and acoustic-phonetic processing deficits ($\sigma_N^2 > 0$) on the model's likelihood function. Finally, Equation 4.18 summarizes the full model for the new BIASES-A from which behavioral data can be simulated, where $p(z_T|V, r_{/ut/})$ is the probability of a /*t*/-response given a stimulus with VOT value *V* from the */*dut*/–*toot* continuum and $p(z_T|V, r_{/uk/})$ is the probability of a /*t*/-response given a stimulus with VOT value *V* from the *duke*–*/*tuk*/ continuum.

Equation 4.18

$$p(z_T|V, r_{/ut/}) = \frac{1}{1 + e^{-\left[\phi \cdot \log\frac{\kappa_{toot}+\alpha}{\alpha} + \log\frac{\left[e^{-\frac{(V-\mu_{/t/})^2}{2(\sigma^2+\sigma_N^2)}} + \varepsilon \cdot \left(e^{-\frac{(V-\mu_{/d/})^2}{2(\sigma^2+\sigma_N^2)}} - e^{-\frac{(V-\mu_{/t/})^2}{2(\sigma^2+\sigma_N^2)}} \right)\right]}{\left[e^{-\frac{(V-\mu_{/d/})^2}{2(\sigma^2+\sigma_N^2)}} + \varepsilon \cdot \left(e^{-\frac{(V-\mu_{/t/})^2}{2(\sigma^2+\sigma_N^2)}} - e^{-\frac{(V-\mu_{/d/})^2}{2(\sigma^2+\sigma_N^2)}} \right)\right]}\right]}}$$

$$p(z_T|V, r_{/uk/}) = \frac{1}{1 + e^{-\left[\phi \cdot \log\frac{\alpha}{\kappa_{duke}+\alpha} + \log\frac{\left[e^{-\frac{(V-\mu_{/t/})^2}{2(\sigma^2+\sigma_N^2)}} + \varepsilon \cdot \left(e^{-\frac{(V-\mu_{/d/})^2}{2(\sigma^2+\sigma_N^2)}} - e^{-\frac{(V-\mu_{/t/})^2}{2(\sigma^2+\sigma_N^2)}} \right)\right]}{\left[e^{-\frac{(V-\mu_{/d/})^2}{2(\sigma^2+\sigma_N^2)}} + \varepsilon \cdot \left(e^{-\frac{(V-\mu_{/t/})^2}{2(\sigma^2+\sigma_N^2)}} - e^{-\frac{(V-\mu_{/d/})^2}{2(\sigma^2+\sigma_N^2)}} \right)\right]}\right]}}$$

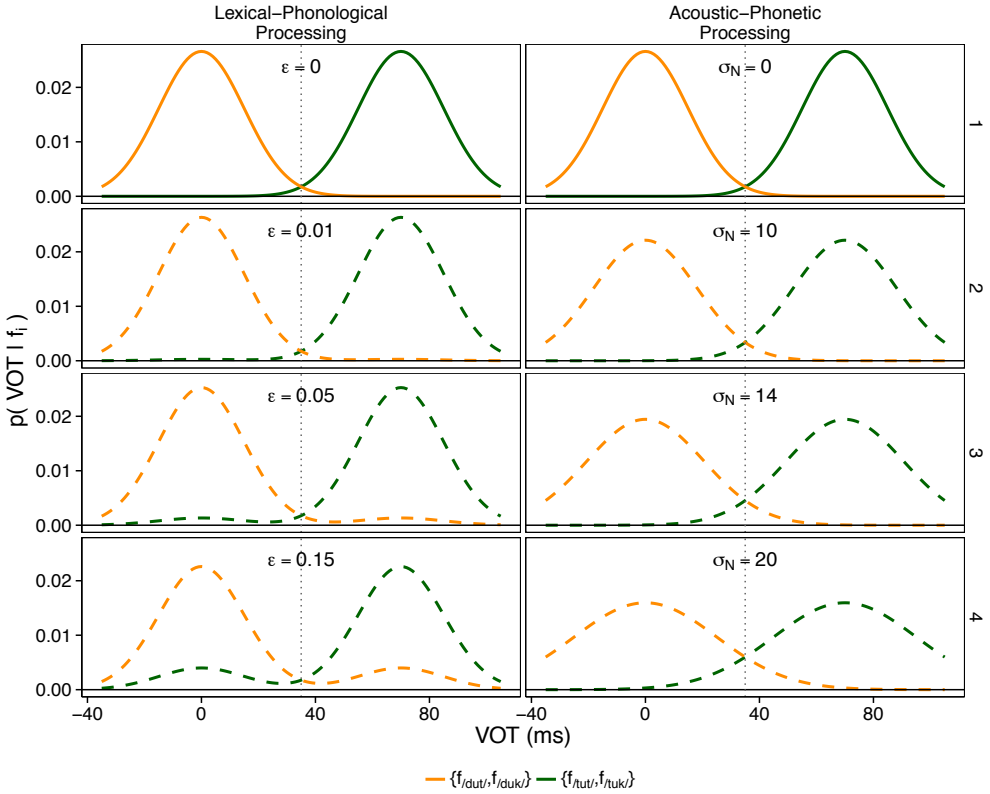

Figure 4.2. Dissociable influences of two bottom-up processing components on speech recognition. Acoustic-phonetic and lexical-phonological processing deficits are modeled as unique information processing transformations that, together, comprise the likelihood function of *BIASES-A*. Virtual lesions to each has unique effects on the probability density functions, $p(V|f_i)$, of each f_i, where each f_i is a lexical candidate (e.g., *toot* vs. **dut*) competing for recognition in a lexical buffer. The influence of acoustic-phonetic processing impairments (modeled as an increase in the category variance, σ_N) is to render a wider range of VOTs "somewhat representative" of each f_i. The influence is assumed to be uniform over all onsets (and therefore over all f_i). On the other hand, the influence of lexical-phonological processing impairments is modeled as an increased chance of "mishearing" the onset due to an increasingly noisy mapping between word-forms and onsets, where the probability of an errant lexical-phonological mapping is given by $\varepsilon = p(/t/|dut) = p(/d/|toot)$. The rate of noisy mappings is assumed to be symmetrical across word-forms: the probability of a word-form whose true onset is /*d*/ (e.g., **dut*, *duke*) being activated when the listener perceives a /*t*/ onset is equal to ε, which is also equal to the probability of a word-form whose true onset is /*t*/ (e.g., *toot*, **tuk*) being activated when the listener perceives a /*d*/ onset. The influence of ε is to increase the bimodality of the density function since BIASES' likelihood is a mixture of Gaussians. Although it is possible for both levels of processing to be impaired (and for BIASES to detect both impairments by implicitly identifying and teasing apart their independent contributions; see results of Experiment 4.1), the simulations presented here only vary one dimension at a time. On the left panels, $\sigma_N = 0$, while on the right $\varepsilon = 0$ (baseline levels).

4.3. Top-Down Effects of Lexical Status on Spoken Word Recognition in Aphasia

Simulation Study 4.1 and Experiment 4.1 were designed to investigate the role of lexical status in spoken word recognition in aphasia. As discussed earlier, the classic finding about top-down lexical effects on phoneme identification is that listeners (who are neurologically healthy) show biases in their labeling of a phonetically ambiguous spoken segment based on the subsequent phonetic material when only one of the two competing candidate labels for the segment would represent a word in the listener's language (Ganong, 1980).

Although the stimuli and task used in Experiment 4.1 and simulated in Simulation Study 4.1 are described in greater detail later on (see **Sections 4.3.2.1.2 - 4.3.2.1.3**), the following represents an overview of key aspects of the Methods. Subjects (including both healthy controls and patients with aphasia) heard tokens from a VOT continuum between /*d*/ and /*t*/ that were immediately followed by one of two rimes (/-*uk*/ or /-*ut*/) and their task was to decide whether the first segment was an exemplar of a /*d*/ or of a /*t*/. Critically, when the segment was followed by /-*uk*/, a /*d*/-response corresponded to a word-response (because *duke* is a word, but */*tuk*/ is not), but when the segment was followed by /-*ut*/, a /*t*/-response corresponded to a word-response (because *toot* is a word, but */*dut*/ is not). In this stimulus set, the presence of a top-down lexical effect would therefore be realized if, for the same ambiguous VOT token, /*t*/-responses were more likely in the /-*ut*/ condition than in the /-*uk*/ condition (Blumstein et al, 1994; Burton, Baum & Blumstein, 1989).

As described earlier, BIASES-A captures this lexical bias by assuming that subjects' prior expectations for the words *duke* and *toot* are stronger than their

expectations for non-word stimuli like */*tuk*/ and */*dut*/. As discussed in Chapter 3, just how strong the lexical bias is appears to be (i.e., how large the top-down lexical effect is) depends on several factors, including the strength of subjects' top-down lexical expectations (captured by the model's prior) and the degree of bottom-up acoustic ambiguity of a stimulus (captured by the model's likelihood). Consequently, to the extent that patients with aphasia suffer from disruptions to either their lexical-level processing or their bottom-up acoustic/phonetic/phonological processing, signatures of these impairments should be present in their behavioral response patterns. BIASES-A allows us to characterize the signatures associated with different functional linguistic deficits.

To that end, Simulation Study 4.1 examines the expected consequences of disruptions at three levels of processing and Experiment 4.1 assesses the extent to which patients actually do exhibit atypical patterns of top-down effects in their behavioral responses to these stimuli. Ultimately, we can leverage the theoretical framework provided by BIASES-A in order to assess the extent to which the responses of patients with BA and patients with W/CA indicate bottom-up processing deficits, lexical-level processing deficits, or both.

4.3.1. Simulation Study 4.1: Lexical Effects in Aphasia

Simulation Study 4.1 examined the independent contributions of lesions at three different processing levels on the expected size of lexical effects. The results of these simulations are summarized in Figure 4.3. In short, lesions to the prior (ϕ, which controls the weighting of lexical/frequency information) predict atypical patterns in behavioral results that are most notable for the exaggerated (when $\phi > 1$) or diminished (when $\phi < 1$) effect sizes close to the phonetic category boundary (where acoustic information is most

ambiguous). On the other hand, as illustrated in Figure 4.2, lesions to the likelihood ($\varepsilon > 0$ or $\sigma_N > 0$) tend to predict atypicalities with respect to which VOT values are judged to be ambiguous, while the maximum effect size remains relatively unchanged. For lexical-phonological processing deficits, patients tend to mix up endpoint tokens (e.g., by mislabeling, or mishearing, clear exemplars of */*tuk*/ as *duke*) at higher rates. Note that the size expected top-down effects is greater at the /*t*/ end of the VOT continuum than at the /*d*/ end of the continuum. This is the result of the /*t*/ endpoint token that is a word is less frequent than the /*d*/ endpoint token that is a word ($\kappa_{duke} > \kappa_{toot}$), illustrating the interacting roles of the prior and the likelihood during speech integration. Finally, for acoustic-phonetic processing impairments, lesions are expected to induce top-down effects for a wider array of VOTs, just as the addition of noise would cause.

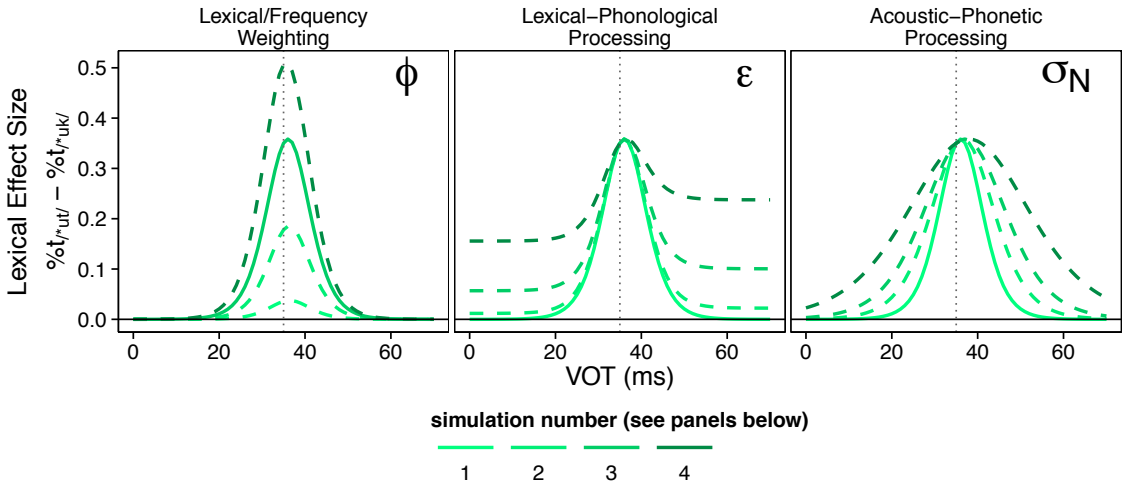

Figure 4.3. Summary of results of Simulation Study 4.1: Effect of manipulating each parameter on the predicted lexical effect size, as a function of VOT. Each curve represents the difference between the posterior probability functions of the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions. In each panel, only the labeled parameter was manipulated; other baseline parameter values ($\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$) were held constant in order to observe the effects of each parameter independently. Solid curves represent the simulation in each panel for which all baseline assumptions were held constant. Each panel summarizes four simulations (i.e., four levels of the relevant parameter for that panel), whose coloration corresponds to the panel number in which that simulation is further detailed in Figures 4.4, 4.5 or 4.6. Coloration darkens from simulation 1-4, because the boundary shift associated with that simulation also increased from simulation 1-4 in each panel. This can be seen in Figures 4.4, 4.5 or 4.6, which show the two conditions' posterior probability curves, as a function of VOT.

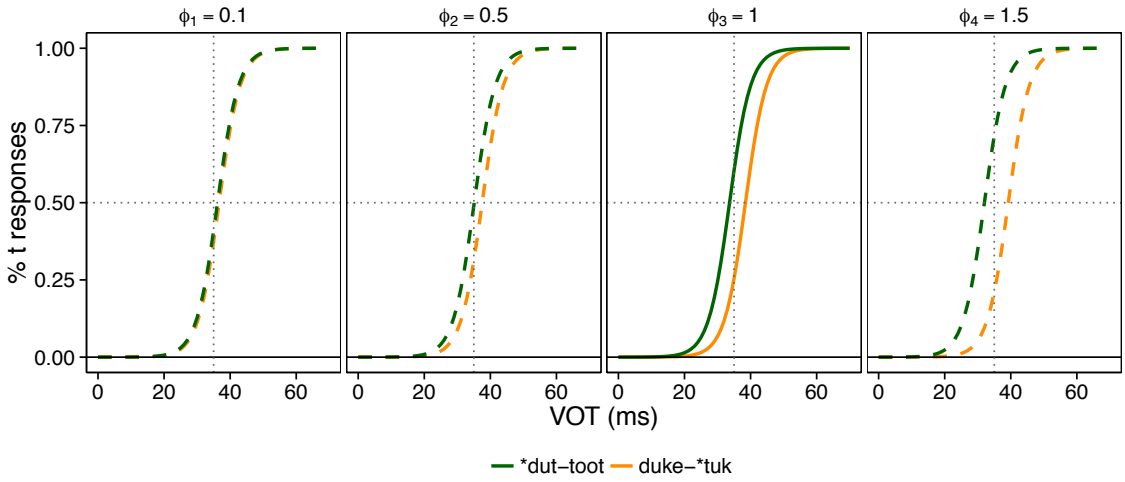

Figure 4.4. Detailed Results of Simulation Study 4.1: Effect of weighting lexical information (ϕ) on expected rate of voiceless (/*t*/) responses, as a function of VOT and rime of the stimulus (/*uk*/ vs. /*ut*/), which corresponded to opposing lexical biases for the initial consonant. The panel with solid lines represents the baseline assumptions ($\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$), and each other panel manipulated only the listed parameter value; all others remained at baseline. The vertical grey line denotes the phoneme category boundary in the simulations (the VOT at which, for an unbiased prior, the posterior probability of /*t*/- and /*d*/-response are equal.

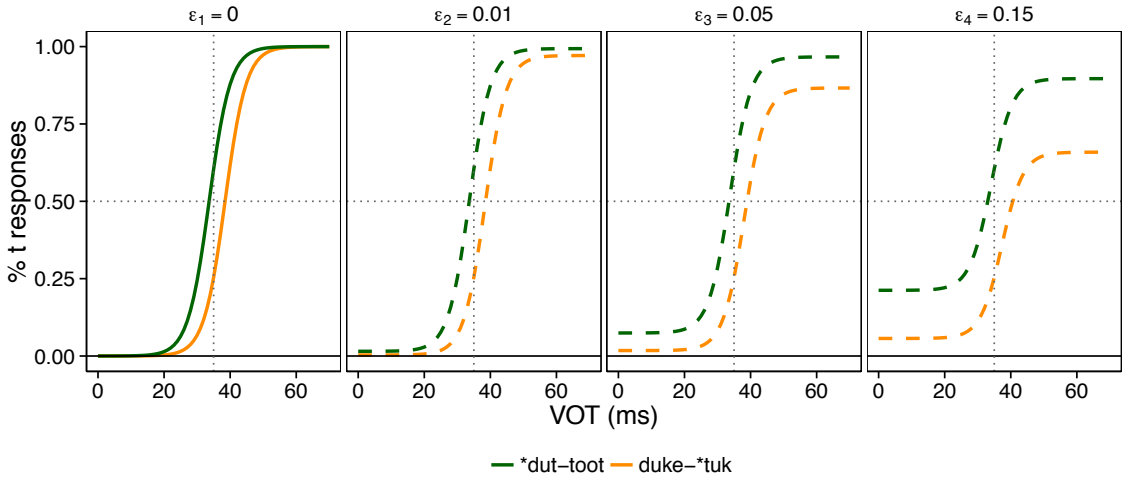

Figure 4.5. Detailed Results of Simulation Study 4.1: Effect of efficacy of phonological processing (ε) on expected rate of voiceless (/*t*/) responses, as a function of VOT and rime of the stimulus (/*uk*/ vs. /*ut*/), which corresponded to opposing lexical biases for the initial consonant. The panel with solid lines represents the baseline assumptions ($\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$), and each other panel manipulated only the listed parameter value; all others remained at baseline. The vertical grey line denotes the phoneme category boundary in the simulations (the VOT at which, for an unbiased prior, the posterior probability of /*t*/- and /*d*/-response are equal.

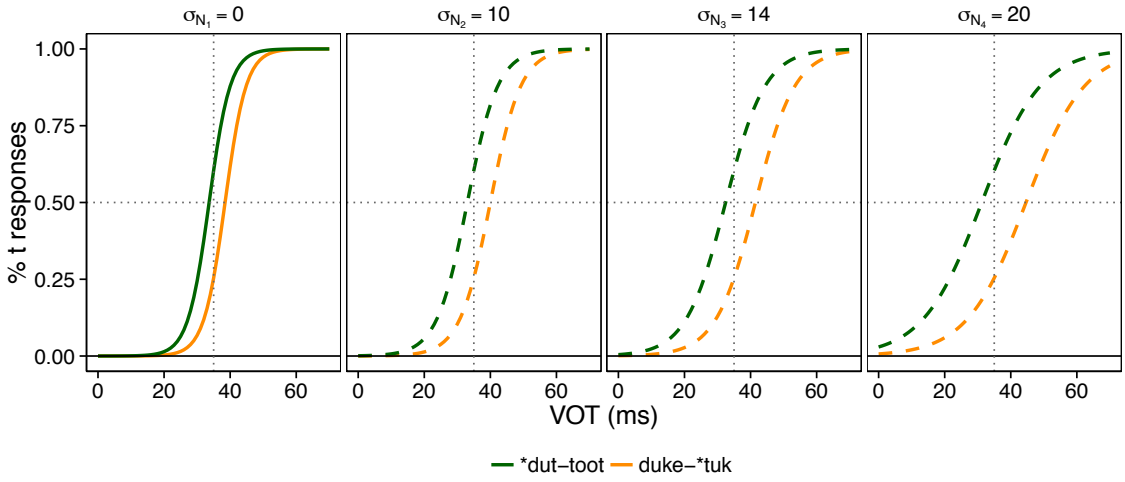

Figure 4.6. Detailed Results of Simulation Study 4.1: Effect of efficacy of acoustic-phonetic processing (σ_N) on expected rate of voiceless (/*t*/) responses, as a function of VOT and rime of the stimulus (/*uk*/ vs. /*ut*/), which corresponded to opposing lexical biases for the initial consonant. The panel with solid lines represents the baseline assumptions ($\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$), and each other panel manipulated only the listed parameter value; all others remained at baseline. The vertical grey line denotes the phoneme category boundary in the simulations (the VOT at which, for an unbiased prior, the posterior probability of /*t*/- and /*d*/-response are equal.

4.3.2. Experiment 4.1: Lexical Effects in Aphasia

Based on the simulations with BIASES-A presented in Simulation Study 4.1, and based on the logic outlined above in **Section 4.2.2** (see Table 4.1 for a summary), it is clear that the Lexical Activation Hypothesis predicts that BAs should exhibit exaggerated lexical effects compared to healthy control subjects and W/CAs should exhibit diminished (or perhaps even undetectable) lexical effects. However, as discussed earlier, both patient groups may also suffer from bottom-up processing deficits. Simulation Study 4.1 further illustrated that bottom-up and lexical-level processing deficits have distinct signatures in the expected behavioral results, suggesting that, by analyzing the fine-grained pattern of behavioral responses from a study examining the lexical effect in patients with aphasia (as compared to healthy listeners) in light of the predictions of

BIASES-A, it may be possible to tease apart the impacts of different functional impairments and infer the nature of the underlying deficits in the BAs and W/CAs.

To that end, the data from Experiment 4.1, described below, were examined in order to evaluate the extent to which these data support the predictions of the Lexical Activation Hypothesis. As previously mentioned, the data for Experiment 4.1 were originally presented by Blumstein and colleagues (1994). Here, we reanalyze the raw data from that study using the model-based approach.

4.3.2.1. Methods

For a detailed description of the participants, stimuli, and procedure of the original study by Blumstein and colleagues (1994), readers should consult that article. However, a summary is provided below.

4.3.2.1.1. Subjects

A total of thirty subjects participated in Experiment 4.1, including 10 young control subjects, 8 age-matched control subjects, 6 patients diagnosed with Broca's aphasia, and 6 patients diagnosed with either Wernicke's or Conduction aphasia.

Ten Brown University students participated, serving as the young control (YC) sample. All reported having normal hearing and being native speakers of English.

Eight right-handed males with a mean age of 64.0 years (minimum: 55; maximum: 75; sd: 6.5) participated, serving as the age-matched control (AMC) sample. All reported having normal hearing and being native speakers of English.

Six patients with Broca's aphasia (mean age: 58.3 years; minimum: 44; maximum: 72; sd: 9.9) participated, comprising the BA sample. Four other patients with Broca's aphasia were originally selected to participate in the study, but were excluded

from the original study after a pre-test because they were unable to accurately identify phonetically unambiguous exemplars of the stimuli included in the study. Patients' clinical diagnoses were determined based on clinical and neurological examinations (including CT scans) and performance on the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983).

Six patients with Wernicke's or Conduction aphasia (mean age: 65.2 years; minimum: 52; maximum: 78; sd: 11.2) participated, comprising the W/CA sample. Four additional patients with Wernicke's aphasia were excluded from the original study after failing the same pre-test administered to the patients in the BA sample, and clinical diagnoses were determined according to the same criteria.

4.3.2.1.2. Stimuli

The stimuli for Experiment 4.1 were comprised of a total of 14 acoustic tokens from two continua that crossed initial consonant voicing with lexical status. In particular, they consisted of 7 tokens from a continuum between a word and a non-word (W-NW continuum; *duke–*tuk*) and 7 acoustic tokens with the same word-initial voice-onset time (VOT) values, but ending in a different final consonant (NW-W continuum; **dut–toot*).

The stimuli were a subset of those used in another previously published study (Burton, Baum & Blumstein, 1989). Burton and colleagues (1989) constructed two 12-step VOT continua. A naturally produced token of *duke* served as the /*d*/ endpoint of the *duke–*tuk* continuum. The other 11 steps of the *duke–*tuk* continuum were constructed by acoustically manipulating this token's waveform, splicing out successively longer portions of the vowel and inserting equal durations of aspiration from the naturally produced **tuk* token. Additionally, the *duke* token's burst was replaced with the **tuk*

token's burst, and the amplitude of the burst varied over the continuum (see Burton et al, 1989). Finally, the tokens of the **dut–toot* continuum was constructed by replacing the final /*k*/ of the tokens from the *duke–*tuk* continuum with the final /*t*/ from a naturally produced token of **dut*, thus ensuring that the W-NW continuum and the NW-W continuum did not differ acoustically except in their final consonant.

Blumstein and colleagues (1994) selected 7 of the 12 stimuli from each VOT continuum, corresponding to two /d/ endpoint tokens (VOTs = 14.7 and 18.7 ms), two /t/ endpoint tokens (VOTs = 55.7 and 60.2 ms), and three phonetically ambiguous tokens with intermediate VOTs (VOTs = 34.2, 37.3, and 41.7 ms).

4.3.2.1.3. Procedure

The seven tokens from each VOT continuum were binaurally presented over headphones to each participant 10 times. Subjects heard stimuli from each continuum in two separate tests separated by a short break (order of presentation of the two continua was counterbalanced across subjects). The 70 trials for a given continuum were randomly ordered and presented in blocks of 10 trials, with sequential blocks separated by a 6-second interval. Trials within a block were separated by a 3-second inter-stimulus interval for young control subjects and a 4-second inter-stimulus interval for all older subjects (age-matched controls and all patients with aphasia). Subjects were instructed to identify the first sound of each stimulus as either "d" or "t" by pressing the appropriately labeled button (counterbalanced across subjects) with their preferred hand as quickly and accurately as possible. Prior to each test, all subjects completed 12 randomly ordered practice trials, including at least one trial with each of the 7 tokens from the continuum being tested and 5 additional trials with randomly selected tokens from that continuum.

Prior to the experiment, all patients with aphasia completed a pretest in which they heard each of the /*d*/ and /*t*/ endpoint stimuli from each continuum (VOTs = 14.7 ms and 60.2 ms) ten times. Only participants who achieved at least 70% accuracy on each of the endpoint VOTs completed the experiment (six participants in each patient group).

4.3.2.2. Results: Statistical Analyses

The results of Experiment 4.1, as originally reported by Blumstein and colleagues (1994) are shown in Figure 4.7. Recall that, to the extent that subjects tend to label stimuli with the same word-initial VOT as beginning with a /*t*/ more often in the **dut*–*toot* continuum than in the *duke*–**tuk* continuum, those results would suggest that subjects are biased towards the word endpoint of each continuum, and such results would, in turn, represent evidence of top-down effects from lexical status on speech recognition.

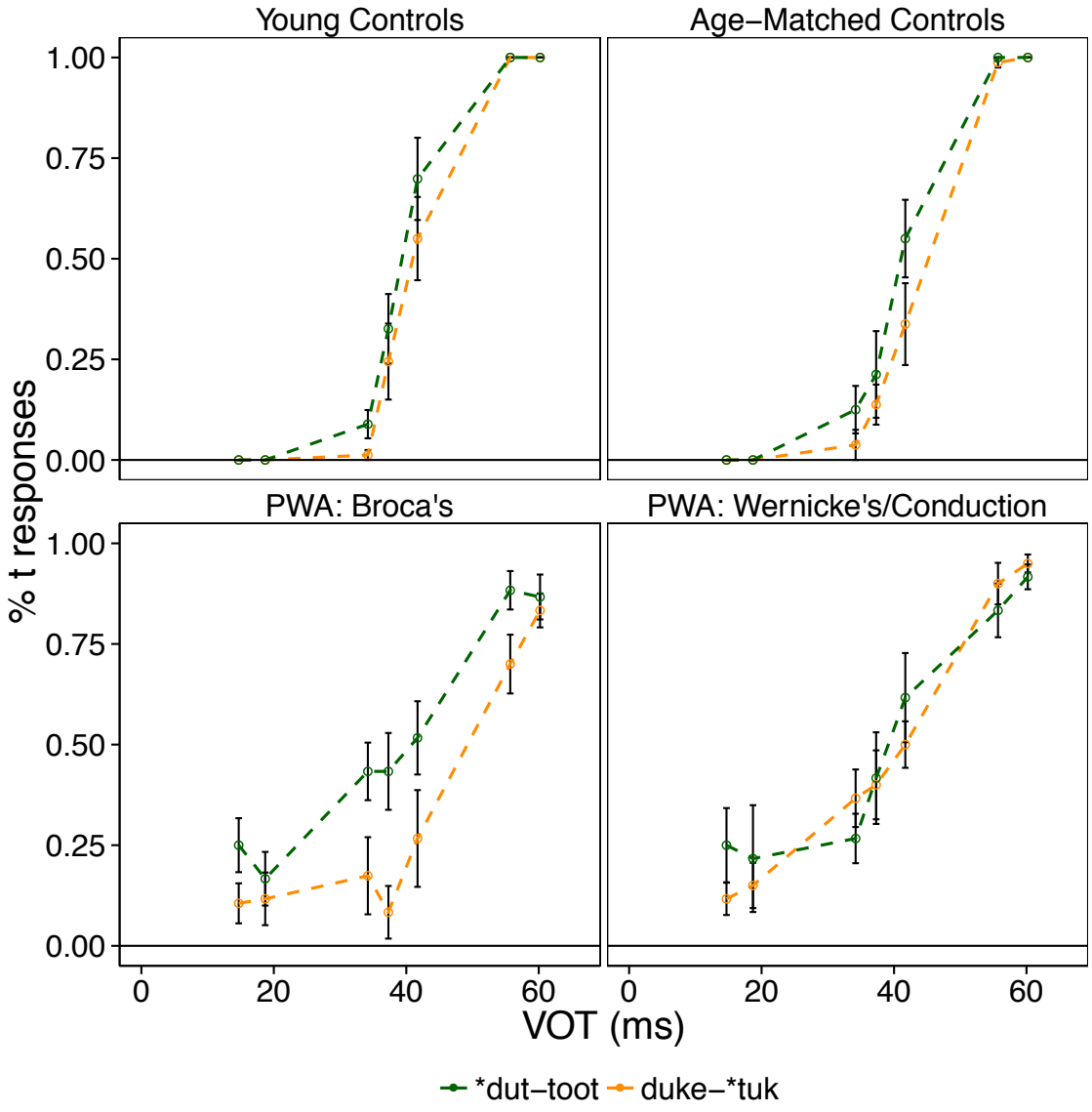

Figure 4.7. Results of Experiment 4.1: for each group, the proportion /*t*/-responses as a function of voice-onset time (VOT) for the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions. Error bars represent by-subject standard error. Results represent reanalysis of raw data from Blumstein et al (1994). PWA = Patients with aphasia.

4.3.2.2.1. Motivation and Interpretation of Logistic Regressions

We reanalyzed the raw data from Experiment 4.1 (that is, the number of /*t*/-responses by each subject, for each VOT value, in each continuum: *duke–*tuk* vs. **dut–toot*) using mixed effects logistic regression (Baayen, Davidson & Bates, 2008; Jaeger,

2008), implemented using the lme4 package (Bates, Maechler, Bolker & Walker, 2014) in *R* (R Core Team, 2014). As noted earlier, the coefficients of a logistic regression relate directly to the underlying parameters of a Bayesian model of speech perception in the context of a two-alternative forced choice task (Feldman et al, 2009, Appendix B; see also Kleinschmidt & Jeager, 2015, Appendix, pp. 200-201). In particular, the theoretical framework defined by BIASES implies the appropriate structure for the logistic regression models, and how significance levels should be interpreted.

Consistent with the theoretical framework provided by BIASES, all analyses reported in this section included independent fixed effects for RIME (β_2) (/-*ut*/ vs. /-*uk*/; or, equivalently, /*t*/-biased stimulus vs. /*d*/-biased stimulus) and for VOT (β_1) (modeled here as a continuous, linear fixed effect). No RIME × VOT interaction term was included, reflecting the principle that the prior and the likelihood are independent sources of information in the Bayesian framework (*cf.* Chapter 2). Any significant main effect of RIME suggests an influence of lexical status. A significant main effect of VOT suggests that subjects' likelihood of making a /*t*/-response depends on the VOT of the stimulus. Thus, these two fixed effects reflect top-down and bottom-up processing, respectively.

Additionally, whenever analyses included subjects from more than one group (e.g., a comparison of Young and Age-Matched Controls or a comparison of all three groups of Elderly subjects to one another), a fixed effect of GROUP was included in the model, along with its interactions with both RIME and VOT. Critically, a significant interaction between RIME and GROUP would reflect reliable differences in top-down processing between the two groups being compared, while a significant interaction between VOT and GROUP would reflect reliable differences in the bottom-up processing

between the two groups. Typically, if two groups differ in their best-fitting intercept coefficient (β_0), it would indicate that subjects from the two groups differed reliably from one another in the locus of their category boundary. However, this result could also be attributable to between-group differences in top-down processing due to the default choice of contrasts used to code factors in these models, a point to which we will return later.

Finally, all analyses also included random by-subject intercepts, thereby allowing subjects to vary in their category boundary around some overall group mean (*cf*. Chapter 3). Prior to analysis, VOT was centered (mean = 0) and RIME was deviation-coded (contrasts: -0.5/0.5 for /-*uk*/ and /-*ut*/, respectively). Deviation-coding was also used for the GROUP factor in analyses comparing groups. In comparisons of the two control groups (Young vs. Age-Matched) the older subjects were represented by the positive contrast. In comparisons of the elderly subjects (Age-Matched Controls vs. BAs vs. W/CAs), the GROUP factor was coded using two planned contrasts. These contrasts were selected to be Age-Matched Controls vs. BAs and Age-Matched Controls vs. W/CAs, with Age-Matched Controls being coded as the negative contrast in both cases.

All results are reported in tables that include the best-fitting estimate of each regression coefficient (β), the estimate's standard error (SE), Wald's z statistic for the estimate of that parameter ($|z|$), and the significance level of the statistic (p). Table 4.2 summarizes the theoretical interpretation of each logistic regression coefficient.

Coefficient	Factor	Related Terms in BIASES	Interpretation of Significance
β_0	Intercept	$-b = -\frac{{\mu_1}^2 - {\mu_2}^2}{2\sigma^2}$	reflects estimate of category boundary
β_1	VOT	$g = \frac{\mu_1 - \mu_2}{\sigma^2}$	reliable bottom-up influence of acoustic-phonetic cues on recognition (*likelihood*)
β_2	RIME	$\log\frac{p(f_1)}{p(f_2)}$	reliable top-down influence of lexical status on recognition (*prior*)

Table 4.2. Summary of theoretical interpretations of logistic regression coefficients

4.3.2.2.2. Control Subjects: YCs vs. AMCs

In order to confirm that Experiment 4.1's stimuli elicited a reliable lexical effect for the healthy control subjects, the data from all Young Control subjects (YCs) and all Age-Matched Control subjects (AMCs) were submitted to logistic regression. Unsurprisingly, results (see Table 4.3) revealed contributions of both bottom-up and top-down effects on speech perception. No significant differences between the two control groups emerged, but there was a marginally significant RIME × GROUP interaction, suggesting that the AMCs may be somewhat more influenced by lexical status than YCs.

Coefficient	β	SE	$\lvert z\rvert$	p
β_0	-1.511 (-1.610)	0.286 (0.286)	-5.273 (-5.638)	< 0.001 (< 0.001)
β_1	0.449 (0.452)	0.028 (0.029)	15.958 (15.458)	< 0.001 (< 0.001)
β_2	0.762 (0.916)	0.177 (0.183)	4.314 (5.012)	< 0.001 (< 0.001)
β_0: YC vs. AMC	-0.807 (-0.599)	0.572 (0.570)	-1.412 (-1.052)	0.158 (0.293)
β_1: YC vs. AMC	-0.078 (-0.087)	0.055 (0.057)	-1.418 (-1.512)	0.156 (0.131)
β_2: YC vs. AMC	0.627 (0.313)	0.353 (0.365)	1.777 (0.858)	0.076 (0.391)

Table 4.3. Results of logistic regression analysis of Experiment 4.1 (Blumstein et al, 1994) that included Young and Age-Matched Controls. Shaded boxes indicate statistically significant effects. Statistics in parentheses report the comparable statistic for an identical analysis that excluded one young control subject (see main text). β_0 = intercept (related to phoneme category boundary); β_1 = VOT (related to gain/slope of sigmoid); β_2 = RIME (related to size of the boundary shift introduced by the lexical/frequency bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

Further examination of the potential source of this effect revealed that one of the 10 YC subjects showed large "anti-lexical effects" on all three ambiguous tokens: across

all three tokens, the subject made 79% /*t*/-responses to the *duke–*tuk* continuum, but only 45% /*t*/-responses to the **dut–toot* continuum. When this subject's data are excluded (statistics reported in parentheses in Table 4.3 to facilitate comparison), the overall pattern is the same, but the marginal RIME × GROUP interaction evaporates completely. This suggests that the marginally significant interaction was being driven by a single subject's atypical behavioral pattern. Although it is impossible to know why this subject was so strongly biased in the opposite direction than predicted, the follow-up analysis suggests that this subject is an outlier. Therefore, in the model-based analyses, this subject's anomalous data were excluded in order to prevent group-level parameter estimation from being unduly influenced.

With no evidence that the YCs and AMCs differed substantially in their overall pattern of responses to these data, each group's data were analyzed separately. Results were consistent with the conclusion of the first analysis, showing both bottom-up and top-down influences on speech recognition in both YCs (Table 4.4) and AMCs (Table 4.5). This was true whether or not the atypical YC subject was included in the analysis (see parentheses of Table 4.4), although the effect of lexical status was more reliable when those data were excluded.

Coefficient	β	SE	$\lvert z \rvert$	p
β_0	-1.102 (-1.299)	0.352 (0.346)	-3.126 (-3.751)	0.002 (< 0.001)
β_1	0.486 (0.492)	0.041 (0.044)	11.736 (11.086)	< 0.001 (< 0.001)
β_2	0.447 (0.754)	0.223 (0.242)	2.002 (3.119)	0.045 (0.002)

Table 4.4. Results of logistic regression analysis of Experiment 4.1 (Blumstein et al, 1994) that included only Young Controls. Statistics in parentheses report the same value for an identical analysis that excluded one young control subject (see main text). β_0 = intercept (related to phoneme category boundary); β_1 = VOT (related to gain/slope of sigmoid); β_2 = RIME (related to size of the boundary shift introduced by the lexical/frequency bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

Coefficient	β	SE	$\lvert z \rvert$	p
β_0	-1.926	0.464	-4.155	< 0.001
β_1	0.412	0.038	10.798	< 0.001
β_2	1.083	0.275	3.941	< 0.001

Table 4.5. Results of logistic regression analysis of Experiment 4.1 (Blumstein et al, 1994) that included only Age-Matched Controls. Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary); β_1 = VOT (related to gain/slope of sigmoid); β_2 = RIME (related to size of the boundary shift introduced by the lexical/frequency bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

In general, these effects were consistent with the results reported by Blumstein and colleagues (1994), although, where they found no significant lexical effect in the YCs, the present results did. Given that the data are identical, this is likely due, at least in part, to our use of a more powerful statistical approach: the analyses of Blumstein and colleagues only examined shifts in the estimated phoneme category boundary as a function of lexical status. However, any minor differences between the present results and those originally reported are of little importance to the theoretical interpretation of the data.

4.3.2.2.3. Elderly Subjects: AMCs vs. BAs vs. W/CAs

A logistic regression examined all of the elderly participants, including the AMCs, and patients from both clinically defined groups, BAs and W/CAs. The results of this analysis are shown in Table 4.6. Overall, there was a significant lexical effect on subjects' responses (more /*t*/-responses in the **dut–toot* continuum than in the *duke–***tuk* continuum), and an overall effect of VOT on subjects' responses (more /*t*/-responses to stimuli with longer VOTs).

Coefficient	β	SE	$\|z\|$	p
β_0	-0.794	0.174	-4.554	< 0.001
β_1	0.190	0.012	15.367	< 0.001
β_2	0.736	0.120	6.133	< 0.001
β_0: AMC vs. BA	0.587	0.494	1.189	0.234
β_0: AMC vs. W/CA	1.491	0.494	3.021	0.003
β_1: AMC vs. BA	-0.209	0.026	-8.101	< 0.001
β_1: AMC vs. W/CA	-0.197	0.026	-7.632	< 0.001
β_2: AMC vs. BA	0.672	0.313	2.148	0.032
β_2: AMC vs. W/CA	-1.245	0.309	-4.031	< 0.001

Table 4.6. Results of logistic regression analysis of Experiment 4.1 (Blumstein et al, 1994) that included Age-Matched Controls (AMC), patients with Broca's aphasia (BA), and patients with Wernicke's or Conduction aphasia (W/CA). Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary); β_1 = VOT (related to gain/slope of sigmoid); β_2 = RIME (related to size of the boundary shift introduced by the lexical/frequency bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

However, both BAs and W/CAs differed from AMCs with respect to both of these effects. In particular, the influence of VOT on speech recognition was diminished in each of the patient groups when compared to the controls. Weaker effects of VOT correspond to a shallower slope of the sigmoidal categorization curve; for reference, a shallower slope is also the expected effect of adding Gaussian noise to a stimulus (Feldman et al, 2009). Thus, this pattern in the results indicates that BAs and W/CAs in Experiment 4.1 both exhibited bottom-up perceptual processing deficits relative to AMCs.

On the other hand, the patterns of lexical effects displayed by the BAs and W/CAs are quite different. Results indicated that both patient groups appeared to differ significantly from AMCs in the extent to which rime (i.e., lexical status) influenced speech perception, but in opposite directions. Whereas BAs were more influenced by lexical status compared to AMCs, W/CAs were less influenced than AMCs. Notably, this is precisely the prediction that emerged in the simulations that aimed to specify the

relationship between the Lexical Activation Hypothesis and a probabilistic speech perception model like BIASES.

Finally, W/CAs also differed significantly from AMCs in the best-fitting intercept. This result has two possible interpretations. The simplest interpretation is that W/CAs (but not BAs) have a different overall category boundary compared to AMCs. This would suggest that these patients' underlying phonetic expectations for the VOTs of exemplars of the /*t*/ and/or /*d*/ phoneme categories are different from controls. While this is certainly possible, there is little evidence that patients with aphasia exhibit fundamentally different phonetic category structure from healthy controls. Quite to the contrary, evidence suggests that, while overall performance on phoneme discrimination and categorization tasks is very often impaired in patients with aphasia (including W/CAs), the typical signatures of phonetic category structure are preserved (Blumstein, Tartter, Nigro & Statlender, 1984; Blumstein et al, 1977b), even in patients who present with specifically impaired acoustic-phonetic processing (e.g., Caplan & Aydellott Utman, 1994; Gow & Caplan, 1996).

Alternatively, the apparent shift in W/CAs' phonetic category boundary could also be explained as an artifact of the smaller lexical effects. The decision to fit the logistic regressions with assumed contrasts for the RIME factor that were equally far from zero (/*-uk*/: -0.5; /*-ut*/: 0.5) implied that the strength of the bias towards /*t*/ imposed by the **dut–toot* continuum and the strength of the bias towards /*d*/ in the *duke–*tuk* continuum should be equal. That is, the fit boundary is exactly halfway between the theoretical boundaries for the two VOT continua. However, if the bias created by the smoothed lexical frequency prior towards *duke* is stronger than the bias towards *toot*, as

predicted by BIASES, then the actual phoneme category boundary should tend to be closer to the implicit category boundary of the **dut–toot* continuum. On the other hand, if W/CAs show a smaller effect of lexical status, overall, then the best-fitting category boundary should be closer to the midpoint between the implicit category boundaries of the two continua (or, if there is no effect of the prior and lexical status, both continua should have the same implicit category boundary, which should be the phonetic category boundary ($\chi = \frac{\mu_1+\mu_2}{2}$).

To further examine the pattern of lexical effects in the two patient groups, each group's data were analyzed separately. Results confirmed that BAs (Table 4.7) exhibited a robust influence of lexical status. Moreover, bottom-up cues (VOT) also influenced speech recognition, although the raw effect size was weaker than in both the control groups. Finally, as suggested by Figure 4.7, W/CAs (Table 4.8) showed no evidence of top-down effects from lexical status. It is especially notable that, although their bottom-up perception of the VOT continua was comparable to BAs, showing a significant effect of VOT with a similar effect size, the primary dimension on which these groups differed was in the extent to which lexical-level information influenced speech recognition.

Coefficient	β	SE	$\lvert z \rvert$	p
β_0	-0.493	0.103	-4.768	< 0.001
β_1	0.085	0.006	13.257	< 0.001
β_2	1.057	0.173	6.129	< 0.001

Table 4.7. Results of logistic regression analysis of Experiment 4.1 (Blumstein et al, 1994) that included only patients with Broca's aphasia (BAs). Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary); β_1 = VOT (related to gain/slope of sigmoid); β_2 = RIME (related to size of the boundary shift introduced by the lexical/frequency bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

Coefficient	β	SE	$\lvert z \rvert$	p
β_0	-0.048	0.156	-0.307	0.759
β_1	0.091	0.006	13.998	< 0.001
β_2	0.112	0.167	0.669	0.503

Table 4.8. Results of logistic regression analysis of Experiment 4.1 (Blumstein et al, 1994) that included only patients with Wernicke's or Conduction aphasia (W/CAs).. Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary); β_1 = VOT (related to gain/slope of sigmoid); β_2 = RIME (related to size of the boundary shift introduced by the lexical/frequency bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $\lvert z \rvert$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

4.3.2.2.4. Summary of Results of Statistical Analyses

Figure 4.8 provides an alternate way of visualizing differences in the size of top-down effects from lexical status for each group over the entire continuum. For each subject's responses to each of the seven VOT tokens, we computed the difference in the proportion of /*t*/-responses in the /*t*/-biased continuum (**dut–toot*) and the /*d*/-biased continuum (*duke–***tuk*), and plotted the mean difference (i.e., effect size) for each group at each VOT. In summary, there are at least five tentative conclusions that find support in the statistical analyses presented above. All are also clearly visible in Figure 4.8:

1. Lexical status influences speech categorization in both YCs and AMCs.
2. Those effects only arise at intermediate VOTs; unambiguous speech tokens are consistently and accurately perceived by healthy adults, even when those speech tokens are non-words (e.g., **dut*).
3. The size of top-down lexical effects and their distribution is nearly identical between YCs and AMCs.
4. The mean size of the lexical effect in patients with BA is greater than in control subjects and lexical influences appear over a wider array of the VOT continua.

5. There is no evidence for an influence of lexical status on speech categorization is W/CAs.

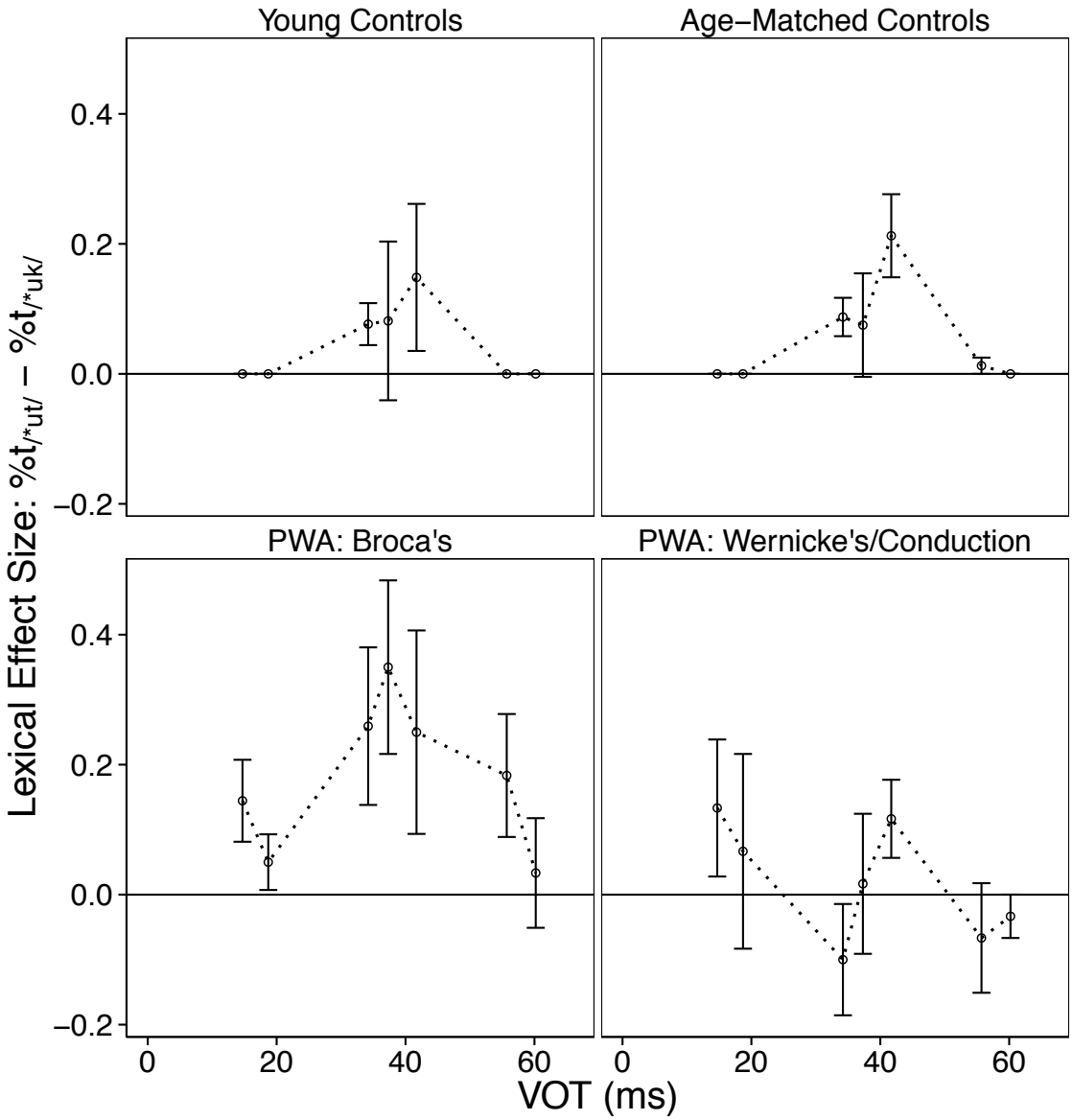

Figure 4.8. Results of Experiment 4.1: Difference between proportion /*t*/-responses in the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions as a function of voice-onset time (VOT), for each group. Error bars represent by-subject standard error. Results represent reanalysis of raw data from Blumstein et al (1994). PWA = Patients with aphasia

4.3.2.3. Results: Model-Based Analyses

The statistical analyses described above provide some evidence for differences in the sizes of top-down effects from lexical status in patients with aphasia compared with

healthy controls. Most interestingly, the results suggest that the identification of stimuli by BAs may be more influenced by lexical status than in healthy control subjects, while W/CAs are less influenced by lexical-level information. According to the simulations summarized earlier in Table 4.1, this is precisely the pattern predicted by the Lexical Activation Hypothesis.

4.3.2.3.1. Motivation of Model-Based Analyses

However, despite these intriguing results, the interpretability of the findings is limited by the analytic techniques employed. Recall that in Chapter 3, it was shown that the size of a boundary shift depends on many factors. Teasing apart competing explanations is not always straightforward. For instance, consider Figures 4.4 and 4.6 in Simulation Study 4.1: in one simulation, manipulating the strength of the bias modulates the size of the boundary shift, while in the other simulation, manipulation the efficacy of acoustic-phonetic processing also modulates the size of the boundary shift. Only one of these parameters is associated with lexical-level processing, so it would be a mistake to conclude from the presence of a larger boundary shift in one group that the difference is attributable to lexical-level processing deficits.

While logistic regression models are much more powerful than comparisons of inferred category boundaries, even these models have other shortcomings. Foremost among these is the relative inflexibility of using generalized linear models (e.g., logistic regression). Such models require a number of assumptions that are not necessarily appropriate for the present work. For instance, consider the influence of ε on expected categorization behavior (see Figure 4.5). The implementation of a fitting procedure for data with asymptotes other than 0 and 1 is not straightforward (Wichmann & Hill, 2001),

but these asymptotes are a direct prediction of our model if subjects suffer from lexical-phonological processing impairments. Furthermore, as with the model intercept that differed between AMCs and W/CAs above (see Table 4.6), some coefficients in logistic regression analyses can be influenced by *both* top-down and bottom-up factors. This fact makes it difficult to isolate the differences caused by lexical-level deficits and those caused by bottom-up processing impairments. Moreover, since multiple unique parameters associated with bottom-up processing dynamics (here, acoustic-phonetic and lexical-phonological processing) are lumped together and expected to influence the same coefficients in a regression model, it is virtually impossible to recover the independent influences of different bottom-up information sources.

Fortunately, model-based Bayesian data analysis makes is possible to explicitly evaluate the independent contributions of multiple interacting model parameters to the observed behavioral data. Such an analytic approach is more theoretically informed, more flexible, more powerful and, ultimately, yields more informative results. Rather than attempting to interpret the relationship between model parameters and regression coefficients (as in Table 4.2), this approach directly models the parameters and processes of interest. Another advantage, especially in situations such as the current one, when data is limited, is that the Bayesian data analysis approach allows a researcher to explicitly choose which parameters should be shared or different between groups or individuals.

4.3.2.3.2. Key Results of Model-Based Analyses

Table 4.9 provides a summary of the posterior distributions of the parameters that were fit in the present analysis (i.e., the "best-fitting" model parameters).

	Mean	SD	95% HDI min	95% HDI max
α	1.25	0.27	0.78	1.82
σ^2	175.28	10.74	154.18	196.59
μ_D	8.04	0.22	7.63	8.50
BAs: σ_N	148.93	82.52	13.64	310.09
W/CAs: σ_N	154.00	67.59	35.69	295.54
BAs: ε	0.14	0.03	0.09	0.19
W/CAs: ε	0.13	0.02	0.09	0.17
BAs: ϕ	1.36	0.30	0.77	1.94
W/CAs: ϕ	-0.12	0.20	-0.51	0.25

Table 4.9. Summary statistics of posterior distributions of Bayesian data analysis of Experiment 4.1 (Blumstein et al, 1994). HDI = highest density interval.

First, it is worth noting that the mean VOT of the /*d*/ onset (μ_D) was estimated to be approximately 8 ms, which is very close to the 5 ms value reported in the seminal analysis of VOTs in English by Lisker and Abramson (1964). This suggests that subjects treated the stimuli like real speech and their responses indicated a category boundary in the typical range.

Of critical interest was the extent to which behavioral responses of BAs and W/CAs might reflect either bottom-up or top-down processing deficits (or both) compared to healthy adults. As suggested by the statistical analyses reported earlier, the model-based analysis provided strong evidence that both patient groups exhibit bottom-up and top-down impairments, but the model-based analyses offer a more detailed picture of the specific deficits underlying abnormal response patterns.

Considering those parameters associated with the efficacy of bottom-up processing first, the results suggest that BAs and W/CAs both suffer from acoustic-phonetic processing deficits, as well as from lexical-phonological deficits. Notably, for both groups, the severity of these bottom-up processing deficits is quite similar in total magnitude (at the group level).

However, when it comes to lexical processing deficits, W/CAs showed significant deficits compared to AMCs (and YCs). As discussed earlier, their responses indicated a weaker influence of lexical status (i.e., frequency) information on phonetic speech categorization decisions. Meanwhile, BAs' lexical processing deficits were trending in the opposite direction. Among these patients, there was a tendency to weight lexical-level cues more heavily than controls (and much more than W/CAs[8]), as predicted by the Lexical Activation Hypothesis.

It is important to note that the simulations in Simulation Study 4.1 focused on illustrating the expected independent impacts of virtual lesions to each processing level of the computational model. The exploratory simulations did suggest the existence of distinct, independent behavioral signatures of bottom-up and lexical-level processing impairments, and the results of the model-based Bayesian data analysis presented here further suggest that the subtle, fine-grained effects of each parameter could be distinguished from one another in the data. However, a more powerful demonstration of this result would be a direct illustration that simulating new behavioral data from a model with the recovered parameter estimates for each group or subjects produced similar patterns of results as the original data. This technique is referred to as a posterior predictive check (PPC), and PPCs can also used to evaluate whether or not a model is sufficient to capture all of the key aspects of the behavioral data.

To further evaluate the ability of the model to fit the data, a posterior predictive check (PPC) was performed. 100 random samples were selected from the joint posterior distribution of the model, and parameter values for a given sample were set to the

[8] Subtracting the posterior chains' samples (BA-W/CA) gives a 95% HDI of [0.26, 3.54], confirming the behavioral divergence on this task for these two clinically defined groups.

sampled value in the corresponding Markov chain. For each sample we simulated data from the model, and we ran all of the statistical analyses reported in **Section 4.3.2.2** on the simulated data. This yielded 100 samples of each of 6 statistical analyses. For each logistic regression coefficient in each statistical test, we computed the mean coefficient estimate ($\boldsymbol{\beta}$) and we determined how many of the statistical tests reached significance at the 0.05 level. To the extent that statistical tests on new, generated data give similar inferences as the same statistical tests on the original data, it would suggest that the model from which the data were generated captures some fundamental aspects of the generative model underlying the psychological processes giving way to the relevant empirical data.

The results are shown in Table 4.10. In general, the PPCs' coefficient estimates and pattern of significances were consistent with the statistics of the original experimental data. Overall, these results suggest that the posterior accurately captured the critical aspects of and patterns in the original data. Figures 4.9 and 4.10 superimpose the results of the PPC onto the original experimental data shown in Figures 4.7 and 4.8.

Experiment 4.1: Lexical Effect		Results: Experiment 4.1		Results: PPC	
logistic regression	**coefficient**	***β***	***p***	**mean *β***	**% sims *p<.05***
Control Subjects	β_0	-1.610	< 0.001	-1.349	100
	β_1	0.452	< 0.001	0.380	100
	β_2	0.916	< 0.001	0.786	96
	β_0: YC vs. AMC	-0.599	0.293	0.006	13
	β_1: YC vs. AMC	-0.087	0.131	0.006	14
	β_2: YC vs. AMC	0.313	0.391	0.033	15
Elderly Subjects	β_0	-0.794	< 0.001	-0.697	100
	β_1	0.190	< 0.001	0.185	100
	β_2	0.736	< 0.001	0.582	99
	β_0: AMC vs. BA	0.587	0.234	0.378	48
	β_0: AMC vs. W/CA	1.491	0.003	0.921	99
	β_1: AMC vs. BA	-0.209	< 0.001	-0.203	100
	β_1: AMC vs. W/CA	-0.197	< 0.001	-0.194	100
	β_2: AMC vs. BA	0.672	0.032	0.905	68
	β_2: AMC vs. W/CA	-1.245	< 0.001	-1.347	91
YCs	β_0	-1.299	< 0.001	-1.352	100
	β_1	0.492	< 0.001	0.377	100
	β_2	0.754	0.002	0.770	84
AMCs	β_0	-1.926	< 0.001	-1.346	100
	β_1	0.412	< 0.001	0.383	100
	β_2	1.083	< 0.001	0.803	92
BAs	β_0	-0.493	< 0.001	-0.508	100
	β_1	0.085	< 0.001	0.083	100
	β_2	1.057	< 0.001	1.034	98
W/CAs	β_0	-0.048	0.759	-0.236	45
	β_1	0.091	< 0.001	0.087	100
	β_2	0.112	0.503	-0.092	11

Table 4.10. Summary of the results of a Posterior Predictive Check (PPC) examining the reliability of the model fit to data from Experiment 4.1 (Blumstein et al, 1994).

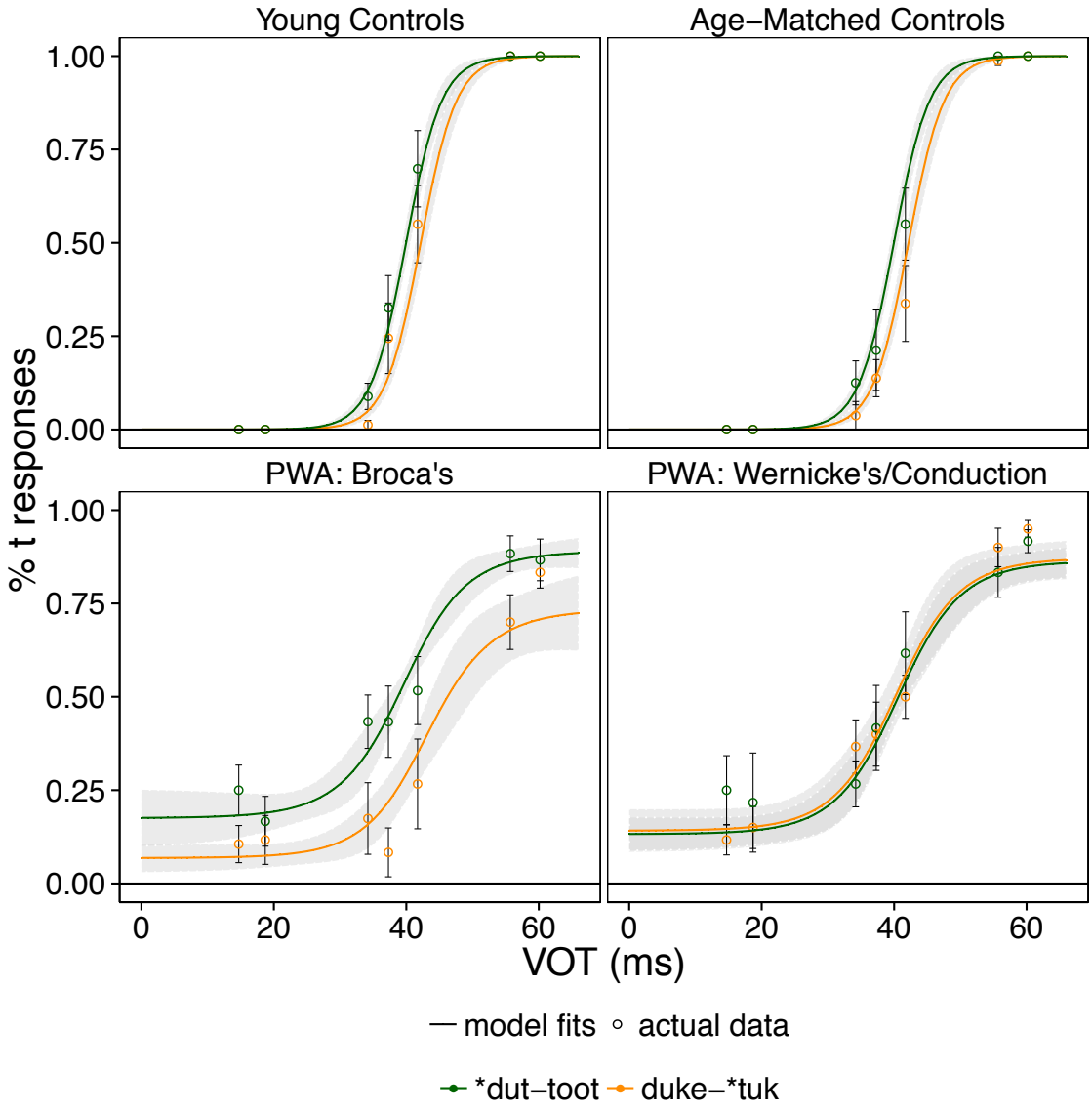

Figure 4.9. Results of Experiment 4.1 (data points; *cf*. Figure 4.7) with superimposed model fits (solid lines). For each group (panel), two curves display the two sigmoidal posterior probability functions of the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions. Young and Age-Matched Controls were fit together. Points indicate proportion /*t*/-responses in the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions for each VOT, for each group. Error bars represent by-subject standard error**.** PWA = Patients with aphasia.

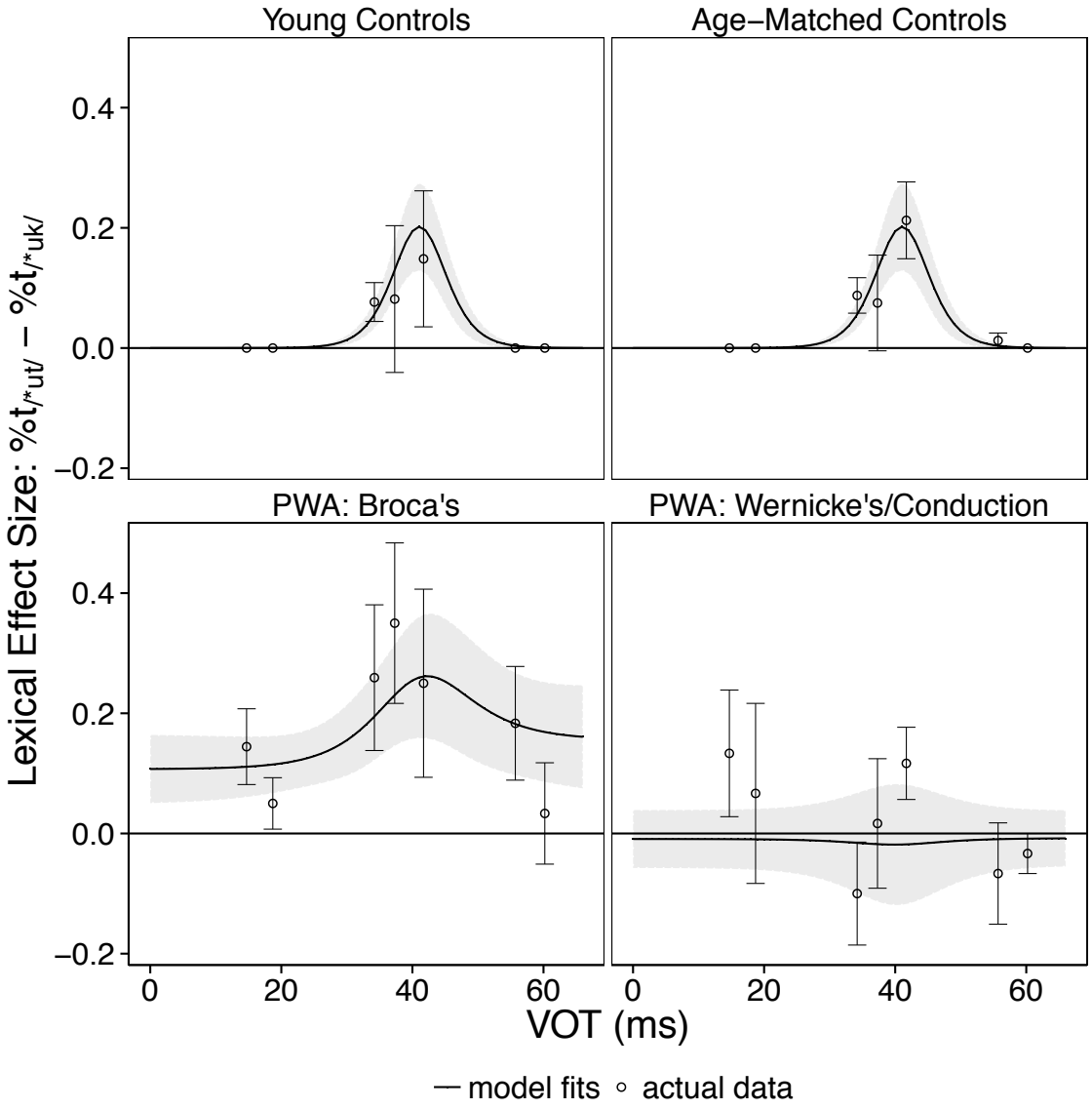

Figure 4.10. Results of Experiment 4.1 (data points; *cf*. Figure 4.8) with superimposed model fits (solid lines). For each group (panel), the curve represents the difference between two sigmoidal posterior probability functions of the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions (*cf*. Figure 4.8). Young and Age-Matched Controls were fit together. Points indicate difference between proportion /*t*/-responses between the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions as a function of VOT, for each group. Error bars represent by-subject standard error**.** PWA = Patients with aphasia.

4.3.2.4. General Discussion of Results of Experiment 4.1

Together with the results of the statistical analyses reported in **Section 4.3.2.2**, the results of the model-based analyses in **Section 4.3.2.3** provide evidence for the

diminished influence of lexical status on the phoneme categorization decisions of patients with W/CA, and (although somewhat less clear) the results may also suggest greater influence of lexical status on the phoneme categorization decisions of patients with BA. These conclusions would be consistent with the predictions of the Lexical Activation Hypothesis. At the same time, the analyses also point towards bottom-up processing impairments (at both the acoustic-phonetic and lexical-phonological levels) in both groups of patients, a finding that is in line with a great deal of work on speech perception in patients with aphasia (Baker, Blumstein & Goodglass, 1981; Basso et al, 1977; Blumstein et al, 1977a, 1977b, 1984; Carpenter & Rutherford, 1973; Jauhiainen & Nuutila, 1977; Leeper, Shewan & Booth, 1986; Metz-Lutz, 1992; Miceli et al, 1978, 1980; Sasanuma et al, 1976; Utman et al, 2001; Yeni-Komshian & Lafontaine, 1983).

At least two other general methodological conclusions also warrant mention. For one, regardless of patient classification, all the patients have a constellation of deficits ranging from acoustic-phonetic to lexical-phonological to lexical-level processing deficits. Although many standard statistical techniques are more limited in the kinds of data they can model and in the kinds of inferences they allow us to draw, BIASES (and in particular BIASES-A) and hierarchical Bayesian data analysis techniques provide a powerful and principled framework for teasing apart subtle differences in the expected influence of different model parameters on subjects' response patterns.

Secondly, another important conclusion is that ignoring patients' clinical classification (and lumping all patients with aphasia into a single group; e.g., "patients with aphasia", or PWA) would preclude us from observing these divergent patterns. To see this, consider Figures 4.11 and 4.12, which merge the two patient groups into one, as

compared to Figures 4.7 and 4.8. It is immediately clear that bottom-up processing deficits are implicated in the broad PWA group, but the opposing lexical-level processing impairments in the two patient groups essentially cancel each other out. Consequently, ignoring clinically relevant classifications could threaten to mask the existence of any lexical-processing deficits at all.

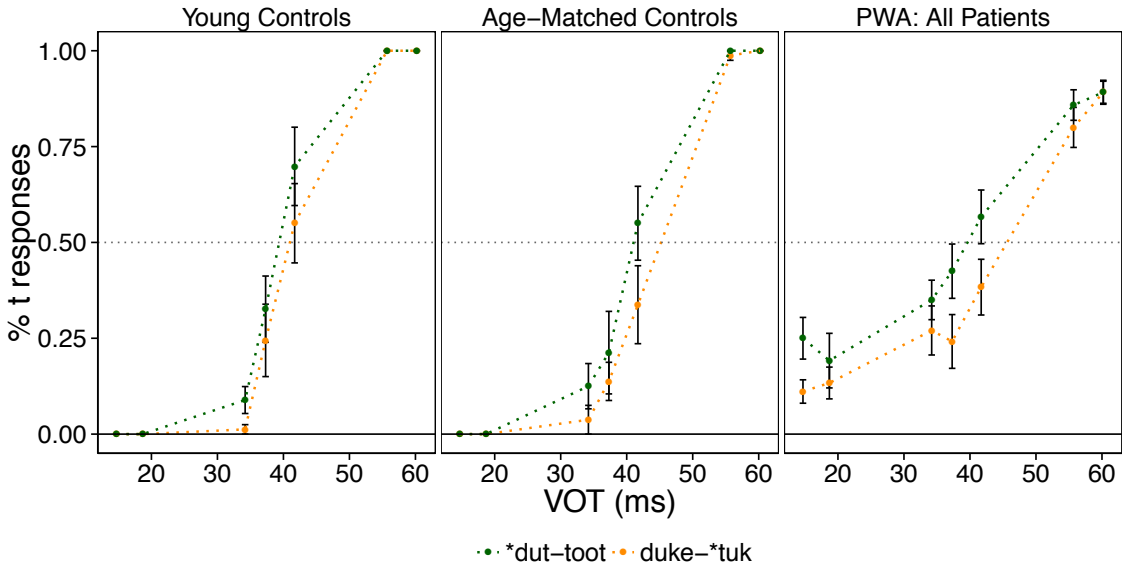

Figure 4.11. Results of Experiment 4.1, merging clinically defined patient groups (BAs and W/CAs) into one single group (PWA = Patients with aphasia). For each group, the proportion of /*t*/-responses as a function of voice-onset time (VOT) for the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions. Error bars represent by-subject standard error. Results represent reanalysis of raw data from Blumstein et al (1994).

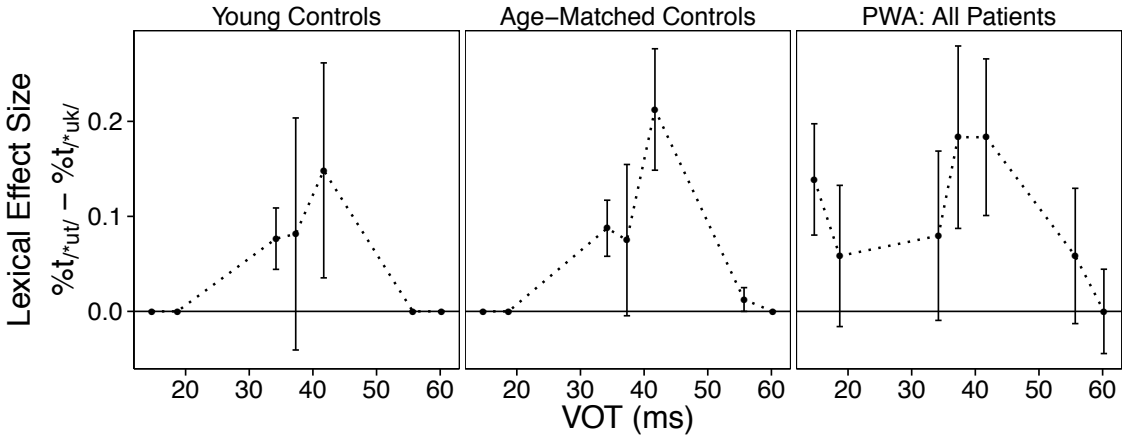

Figure 4.12. Results of Experiment 4.1, merging clinically defined patient groups (BAs and W/CAs) into one single group (PWA = Patients with aphasia). Difference between proportion /*t*/-responses in the /**ut*/ (/*t*/-biased) and /**uk*/ (/*d*/-biased) conditions as a function of voice-onset time (VOT), for each group. Error bars represent by-subject standard error. Results represent reanalysis of raw data from Blumstein et al (1994).

4.4. Top-Down Effects of Sentence Context on Spoken Word Recognition in Aphasia

Broadly, the results of Experiment 4.1 provide evidence that spoken word recognition in patients with aphasia is affected by multiple functional linguistic deficits, including a deficit at the level of lexical processing, as well as deficits in bottom-up processing of the speech signal, and that those deficits (and their consequent effects) differ as a function of clinical diagnosis. However, unlike the stimuli in Experiment 4.1, everyday speech rarely features words uttered in isolation, and it is important to note that, in individuals *without* aphasia, linguistic context has consistently been shown to impact recognition of acoustically ambiguous words. Words that are unintelligible when presented in isolation can often be identified in context (Lieberman, 1963; Pickett & Pollack, 1963; Hunnicutt, 1985; Fowler & Housum, 1987). Furthermore, as discussed at length in Chapters 1-3, stimuli that (when presented in isolation) are perceived as ambiguous between two possible words (e.g., between *bay* and *pay*) tend (when presented in sentences) to be perceived as whichever word is more congruent with a

preceding context (e.g., as *bay* after sentences like *He hated the...* but as *pay* after sentences like *He hated to...*) (Fox & Blumstein, in press; see also Borsky et al, 1998; Connine, 1987; Garnes & Bond, 1976; Guediche et al, 2013; Miller et al, 1984; Rohde & Ettlinger, 2012; Tuinman et al, 2014; van Alphen & McQueen, 2001).

Importantly, most work examining lexical access impairments in aphasia has examined the recognition of isolated words (but see, e.g., Friederici, 1983; Baum, 2001). It remains unclear to what extent spoken word recognition processes in brain-injured patients with aphasia have access to the same information sources during auditory language processing that have been shown to influence speech perception in healthy subjects. This question is also of special interest because sentential context might, in fact, reduce apparent lexical processing deficits in such patients by providing top-down support for those lexical candidates whose processing could ordinarily be impaired when perceived in isolation (as in Experiment 4.1).

Thus, Simulation Study 4.2 and Experiment 4.2 were designed to explore the nature of top-down processing of words by patients with aphasia when the words are embedded in sentential contexts. In particular, the goal of Simulation Study 4.2 was similar to that of Simulation Study 4.1, but further considered the role of sentential context. That is, Simulation Study 4.2 investigates the expected consequences of disruptions at the three levels of processing considered in Simulation Study 4.1 (acoustic-phonetic processing, lexical-phonological processing, and lexical processing), as well as at the level of contextual integration during auditory sentence processing. As we will show, by simulating disruptions at each level of processing in BIASES-A, it is possible to generate fine-grained quantitative predictions about the expected patterns of top-down

effects in patients with various linguistic deficits.

As with Experiment 4.1, Experiment 4.2 was designed to evaluate the extent to which patients with BA and W/CA actually exhibit atypical patterns of top-down effects in their behavioral responses to stimuli embedded in sentences that support one or the other lexical candidate. The stimuli and task employed in Experiment 4.2 (detailed in **Sections 4.4.2.1.2 – 4.4.2.1.3**) resembled the stimuli and task described in Chapters 1-3: subjects (including both healthy controls and patients with aphasia) heard tokens from a VOT continuum between *bay* and *pay* after noun-biasing and verb-biasing sentence contexts (e.g., *He hated the...* vs. *He hated to...*) and their task was to decide whether the last word of each sentence was *bay* or *pay*. Applying the theoretical lens represented by BIASES-A, we submitted these data to a model-based analysis in order to assess the extent to which the responses of patients with BA and patients with W/CA provide evidence for bottom-up processing deficits, lexical-level impairments, deficits affecting the integration of cues from a preceding sentence context, or some combination thereof.

4.4.1. Joint Modeling Contextual and Lexical Effects on Word Recognition

In order to model potential deficits at both lexical and contextual levels of processing and their independent effects on spoken word recognition, one addition was made to BIASES-A. The only difference in the mathematical formulation of BIASES-A was to allow context, *C*, to influence subjects' responses. This was implicit in the original formulation of BIASES-A, because (as described in Chapter 2; see Equation 2.5) lexical frequency is equal to the total number of times the word appears after any context. However, since Experiment 4.1 did not involve any sentential contexts preceding the target stimulus, there could be no influence of context. In order to model the task

examined in Experiment 4.2, though, it was critical to incorporate into BIASES-A both (1) a parameter than can model lexical-level impairments, and (2) a parameter that can model contextual integration impairments. To do so, the form of BIASES-A was updated (Equation 4.19):

Equation 4.19

$$p(f_i|C,V,R) = \frac{p(f_i|C)p(V,R|f_i,C)}{\sum_{j=1}^{N_f} p(f_j|C)p(V,R|f_j,C)}$$

In the updated model, *BIASES-A*, upon perceiving a monosyllabic stimulus and the preceding context (here, limited to the function words *to* vs. *the*), Bayes' rule gives the probability of recognizing a candidate word-form, f_i, given the context, C, the initial segment's voice-onset time, V, and the stimulus's rime, R. As described earlier, we assume that a subject's task is to identify the word-form of a stimulus. Equation 4.19's prior term can be expanded according to Bayes' rule (Equation 4.20).

Equation 4.20

$$p(f_i|C) \propto p(C|f_i)p(f_i)$$

Put simply, Equation 4.20 states that the prior probability of a candidate word-form following context C is proportional to the product of the lexical frequency of the word-form, $p(f_i)$, and $p(C|f_i)$, a term related to the proportion of times the word f_i follows C compared with any other preceding context. That is, $p(C|f_i)$ will be high if, when f_i occurs in a sentence, it usually occurs after C. For instance, the word *Francisco* almost always occurs after *San*, so $p(C = San|f_i = Francisco)$ is high (even though many other cities have names beginning with *San*). On the other hand, $p(C|f_i)$ might be low in two situations: (1) if, when f_i occurs in sentences, it usually occurs after something other than C (e.g., $p(C = Sasha|f_i = Obama) \ll p(C = Barack|f_i = Obama)$), or (2)

if f_i occurs in many contexts such that the occurrence of f_i is not specific to context C (e.g., $p(C|f_i = Smith)$. Thus, Equation 4.20's manipulation of $p(f_i|C)$ includes a term associated with contextual integration and a term associated with lexical-level (frequency) information. The values for $p(C|f_i)$ were estimated from the Google n-grams corpus (Michel et al, 2010) and were smoothed as described in Chapters 2-3.

The same basic assumptions about $p(V, R|f_i, C)$, the likelihood function of BIASES-A described earlier, were maintained here including: that the phonological form of a monosyllabic stimulus is composed of an onset and a rime, that the onset and the rime are conditionally independent cues to the phonological form of the stimulus, that rimes are deterministically related to word-forms, that rimes are consistently perceived accurately, that VOT is the only acoustic cue to the identity of the onset of the stimulus, that the VOTs of acoustic realizations of a given onset follow normal distributions with equal variance for all onsets, and that the distribution of VOTs conditionally independent of lexical or higher-level information given the identity of the onset. Simplifying Equation 4.19 accordingly and applying the straightforward algebraic manipulations described in Chapter 2 (see Equations 2.6-2.7) yields Equation 4.21 (f_1 = *pay*; f_2 = *bay*).

Equation 4.21

$$p(f_1|C,V,R) = \frac{1}{1 + e^{-\left[\log\frac{p(C|f_1)}{p(C|f_2)} + \log\frac{p(f_1)}{p(f_2)} + \log\frac{p(R|f_1)}{p(R|f_2)} + \log\frac{\sum_{k=1}^{No} p(o_k|f_1)p(V|o_k)}{\sum_{k=1}^{No} p(o_k|f_2)p(V|o_k)}\right]}}$$

In order to model the effects of impairments at the acoustic-phonetic, lexical-phonological, and lexical levels of processing, the same three parameters ($\{\sigma_N^2, \varepsilon, \phi\}$) were included as described earlier. A fourth parameter (ω) was also included to model

the influence of impairments in the integration of a preceding contextual cue during spoken word recognition, as shown in Equations 22-24.

Equation 4.22

$$p(f_1|C,V,R) = \frac{1}{1+e^{-\left[\omega\cdot\log\frac{p(C|f_1)}{p(C|f_2)}+\phi\cdot\log\frac{p(f_1)}{p(f_2)}+\log\frac{p(R|f_1)}{p(R|f_2)}+\log\frac{\sum_{k=1}^{No}p(o_k|f_1)p(V|o_k)}{\sum_{k=1}^{No}p(o_k|f_2)p(V|o_k)}\right]}}$$

Equation 4.23

$$p\left(o_{/p/}\middle|f_i\right) = 1 - p\left(o_{/b/}\middle|f_i\right) = \begin{cases} 1-\varepsilon & f_i = pay \\ \varepsilon & f_i = bay \end{cases}$$

Equation 4.24

$$V|o_k \sim N(\mu_k, \sigma^2 + \sigma_N^2)$$

Finally, in the present task, the rime of the target stimulus was always /*ei*/, allowing for yet another simplification. Equation 4.25 summarizes a full model for BIASES-A that allows for independent estimation of parameters associated with contextual integration and lexical-level processing, where $p\left(z_{pay}\middle|C,V\right)$ is the probability of a *pay*-response given a stimulus with VOT value *V* after context *C*.

Equation 4.25

$$p\left(z_{pay}\middle|C,V\right) = \frac{1}{1+e^{-\left[\omega\cdot\log\frac{p(C|pay)}{p(C|bay)}+\phi\cdot\log\frac{p(pay)}{p(bay)}+\log\frac{\left[e^{-\frac{(V-\mu_{/p/})^2}{2(\sigma^2+\sigma_N^2)}}+\varepsilon\cdot\left(e^{-\frac{(V-\mu_{/b/})^2}{2(\sigma^2+\sigma_N^2)}}-e^{-\frac{(V-\mu_{/p/})^2}{2(\sigma^2+\sigma_N^2)}}\right)\right]}{\left[e^{-\frac{(V-\mu_{/b/})^2}{2(\sigma^2+\sigma_N^2)}}+\varepsilon\cdot\left(e^{-\frac{(V-\mu_{/p/})^2}{2(\sigma^2+\sigma_N^2)}}-e^{-\frac{(V-\mu_{/b/})^2}{2(\sigma^2+\sigma_N^2)}}\right)\right]}\right]}}$$

4.4.2. Simulation Study 4.2: Sentential Context Effects in Aphasia

Simulation Study 4.2 examined the independent contributions of lesions at four different processing levels on the expected size of top-down sentential context effects. The results of these simulations are summarized in Figure 4.13. First, it is worth noting

that, as in Simulation Study 4.1 (see Figure 4.3), lesions to the likelihood function ($\varepsilon > 0$ or $\sigma_N > 0$) can best be characterized as driving changes with respect to the distribution of top-down effects over VOT values, but not in the maximum effect size itself. As was seen with the lexical effect simulations (see Figure 4.6), acoustic-phonetic processing impairments (governed by the parameter σ_N) can be expected to induce top-down effects for a wider array of VOTs (see Figure 4.17). Meanwhile, lexical-phonological processing deficits (governed by the parameter ε) are associated with greater effect sizes for endpoint tokens of the VOT continua (see Figure 4.16), reflecting the bottom-up "mishearing" of acoustically clear exemplars of *bay* and *pay* (*cf*. Figure 4.5; see also Figure 4.2).

As for lesions affecting the weighting of information at the lexical level (governed by the parameter ϕ), recall that this parameter was realized responsible for changes in the maximum effect size in the lexical effect simulations (see Figures 4.3 and 4.4). In the present simulations of sentential context effects, that role is played instead by ω, the parameter responsible for the weighting of contextual information during word recognition (see Figure 4.14; compare the leftmost panels of Figure 4.13 and Figure 4.3). This discrepancy is due to the nature of the two tasks being considered and the notion of effect size in the two studies. The two conditions being directly compared in studies of the lexical effect (as in Simulation Study 4.1) differ as a function of lexical information, the influence of which is affected by varying ϕ; however, the two conditions being directly compared in studies of sentential context effects (as in Simulation Study 4.2) differ as a function of how strongly different words are predicted by the preceding cues (i.e., *the* vs. *to*), the influence of which is affected by varying ω. When contextual cues are weighted more strongly (when $\omega > 1$), the relative fit of the candidates (*bay* vs. *pay*)

with the perceived contextual cue will have a greater influence on subjects' behavioral responses, leading to exaggerated top-down context effects, especially when acoustic information is most ambiguous (i.e., close to the phonetic category boundary). The opposite is predicted when contextual cues are weighted *less* strongly (i.e., when $\omega < 1$): the relative fit of competing candidates with the preceding context will be a less reliable predictor of subjects' responses, which will be reflected in diminished top-down context effects.

Although lesions at the lexical level do predict the same types of effects in an experiment examining the size of top-down effects from sentential context on subjects categorization decisions between two words (like the present study) as they do in studies of the lexical effect, they are still predicted to have an effect on word recognition performance. Specifically, increasing and decreasing the weighting of lexical-level cues (governed by the parameter ϕ) tends to shift the locus (on the VOT continuum) of the maximum effect size (see panel 2 of Figure 4.13 and Figure 4.15). The reason for this lies in the relationship between lexical status and lexical frequency. Recall that the current model essentially treats non-words as "very low frequency words" such that every word that appears in a corpus is more frequent than any non-word, while incorporating a mechanism to allow the word recognition system to perceive stimuli that are not found in the corpus from which lexical frequencies are estimated (see **Section 4.2.2.2**). Because *bay* is a less frequent word than *pay*, increasing the weighting of lexical (i.e., frequency) cues leads to a stronger overall bias toward *pay* responses – independent of the preceding context – while decreasing the weighting of the frequency information will tend to reduce the top-down bias towards *pay* (see Figure 4.15). The overall effect of this parametric

variation of ϕ is that as ϕ increases, the center of the distribution of top-down effects shifts closer to the mean of the VOT distribution for the /*b*/ onset, and as ϕ decreases, the center of the distribution of top-down effects shifts closer to the (unweighted) category boundary between the /*b*/ and /*p*/ onsets' VOT distributions. The stronger the frequency bias (i.e., the higher ϕ becomes), the more susceptible otherwise clear tokens of *bay* are to top-down biasing effects from the sentence context, because *pay* is already highly favored as a response; the weaker the frequency bias (i.e., as ϕ approaches 0), the more top-down effects begin to reflect only the fit between the candidates and the preceding context rather than by the lexical frequency of the candidates themselves.

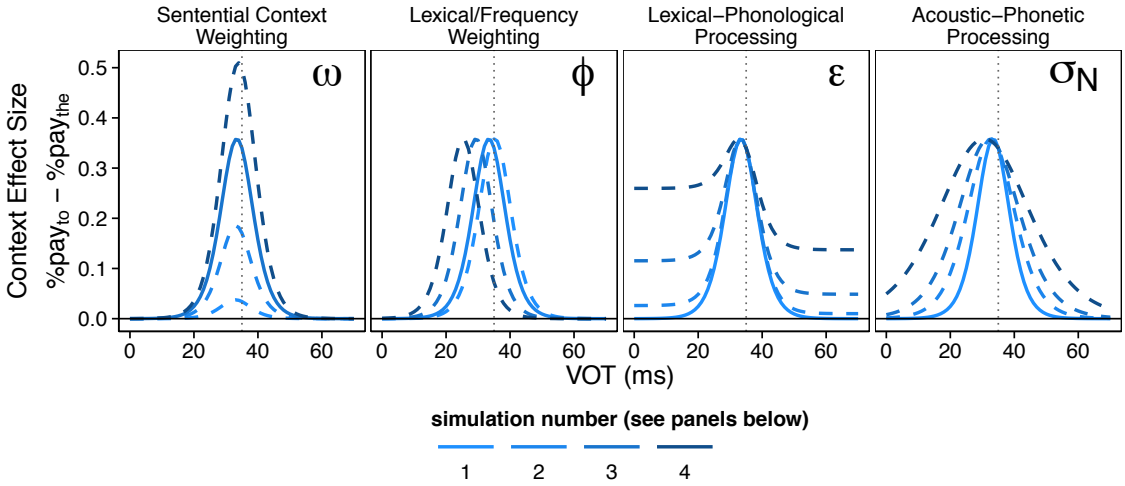

Figure 4.13. Summary of results of Simulation Study 4.2: Effect of manipulating each parameter on the predicted sentential context effect size, as a function of VOT. Each curve represents the difference between the posterior distributions of the *to...* (*pay*-biased) and *the...* (*bay*-biased) conditions. In each panel, only the labeled parameter was manipulated; other baseline parameter values ($\omega = 1$; $\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$) were held constant in order to observe the effects of each parameter independently. Solid curves represent the simulation in each panel for which all baseline assumptions were held constant. Each panel summarizes four simulations (i.e., four levels of the relevant parameter for that panel), whose coloration corresponds to the panel number in which that simulation is further detailed in Figures 4.14, 4.15, 4.16 or 4.17. Coloration darkens from simulation 1-4, because the boundary shift associated with that simulation also increased from simulation 1-4 in each panel. This can be seen in Figures 4.14, 4.15, 4.16 or 4.17, which show the two conditions' posterior probability curves, as a function of VOT.

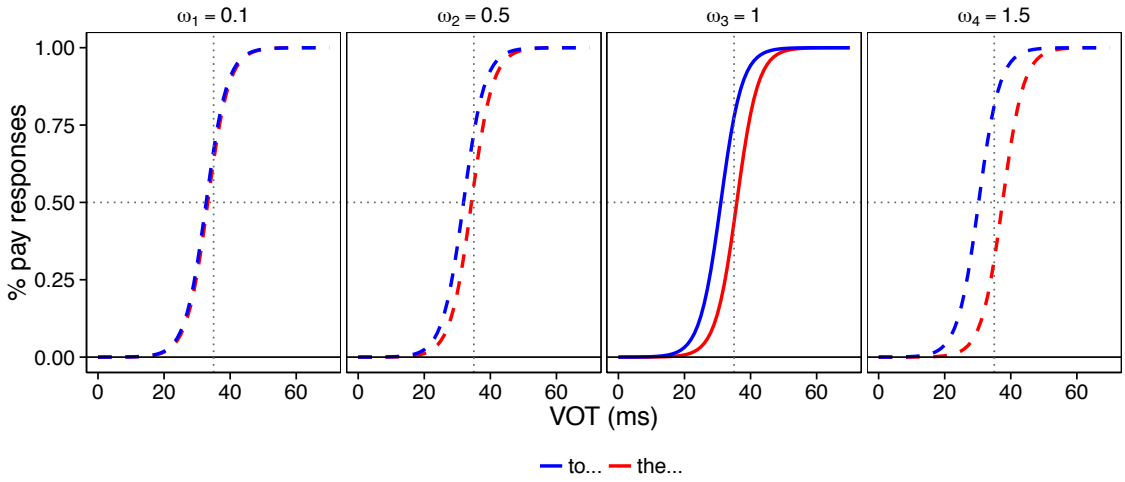

Figure 4.14. Detailed Results of Simulation Study 4.2: Effect of weighting of contextual information (ω) on expected rate of voiceless (*pay*) responses, as a function of VOT and the function word the preceded the stimulus (*to...* vs. *the...*), which corresponded to opposing contextual biases on the initial consonant. The panel with solid lines represents the baseline assumptions ($\omega = 1$; $\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$), and each other panel manipulated only the listed parameter value; all others remained at baseline. The vertical grey line denotes the phoneme category boundary in the simulations (the VOT at which, for an unbiased prior, the posterior probability of *pay*- and *bay*-response are equal.

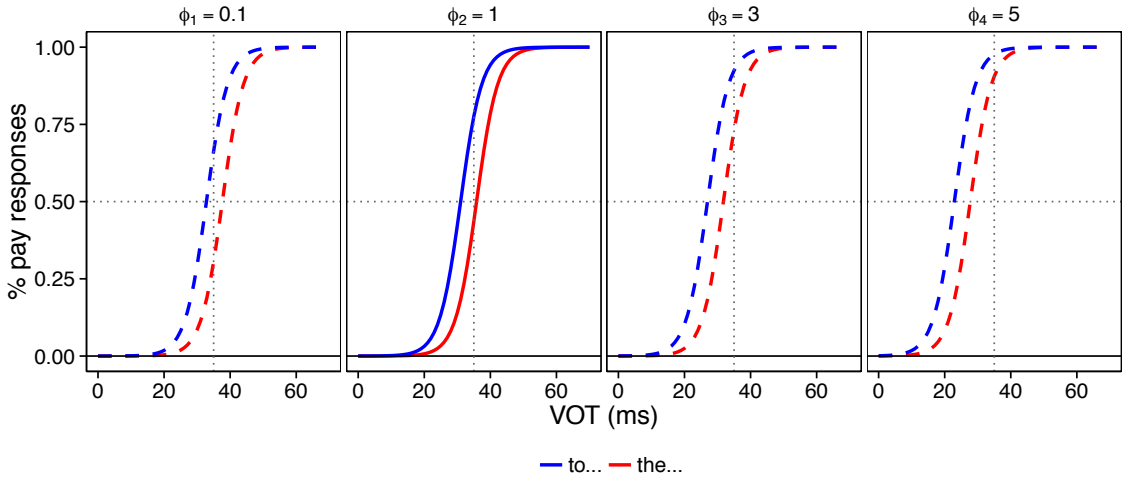

Figure 4.15. Detailed Results of Simulation Study 4.2: Effect of weighting of lexical information (ϕ) on expected rate of voiceless (*pay*) responses, as a function of VOT and the function word the preceded the stimulus (*to...* vs. *the...*), which corresponded to opposing contextual biases on the initial consonant. The panel with solid lines represents the baseline assumptions ($\omega = 1$; $\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$), and each other panel manipulated only the listed parameter value; all others remained at baseline. The vertical grey line denotes the phoneme category boundary in the simulations (the VOT at which, for an unbiased prior, the posterior probability of *pay*- and *bay*-response are equal.

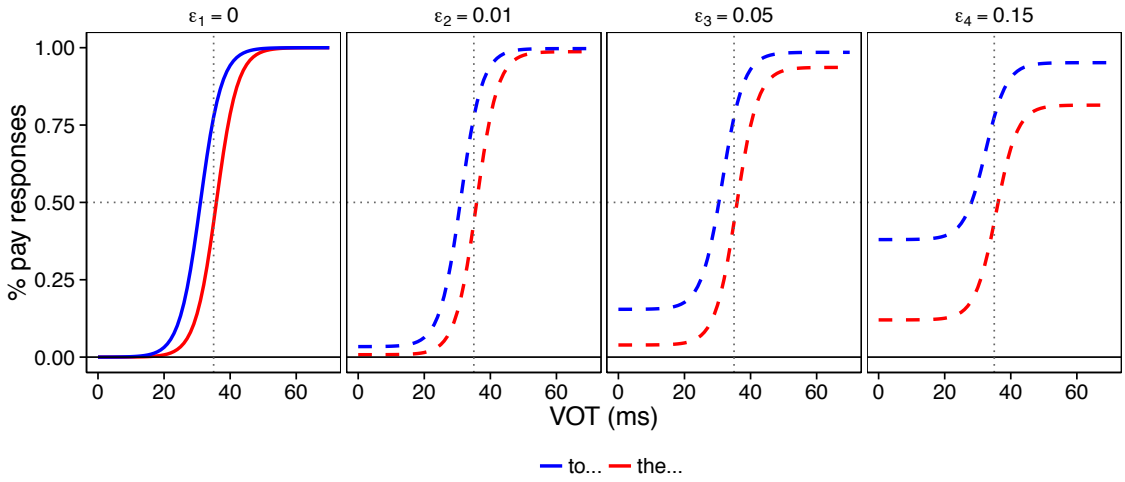

Figure 4.16. Detailed Results of Simulation Study 4.2: Effect of efficacy of phonological processing (ε) on expected rate of voiceless (*pay*) responses, as a function of VOT and the function word the preceded the stimulus (*to...* vs. *the...*), which corresponded to opposing contextual biases on the initial consonant. The panel with solid lines represents the baseline assumptions ($\omega = 1$; $\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$), and each other panel manipulated only the listed parameter value; all others remained at baseline. The vertical grey line denotes the phoneme category boundary in the simulations (the VOT at which, for an unbiased prior, the posterior probability of *pay*- and *bay*-response are equal.

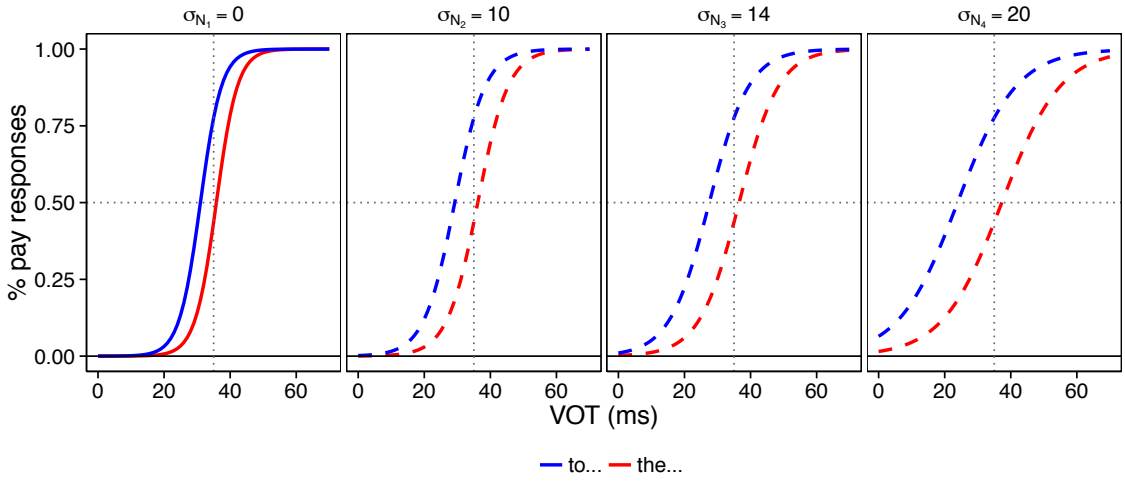

Figure 4.17. Detailed Results of Simulation Study 4.2: Effect of efficacy of acoustic-phonetic processing (σ_N) expected rate of voiceless (*pay*) responses, as a function of VOT and the function word the preceded the stimulus (*to...* vs. *the...*), which corresponded to opposing contextual biases on the initial consonant. The panel with solid lines represents the baseline assumptions ($\omega = 1$; $\phi = 1$; $\varepsilon = 0$; $\sigma_N = 0$), and each other panel manipulated only the listed parameter value; all others remained at baseline. The vertical grey line denotes the phoneme category boundary in the simulations (the VOT at which, for an unbiased prior, the posterior probability of *pay*- and *bay*-response are equal.

4.4.3. Experiment 4.2: Sentential Context Effects in Aphasia

Based on the simulations with BIASES-A presented in Simulation Study 4.2, and based on the logic outlined above in **Section 4.2.2** (see Table 4.1 for a summary), the predictions of the Lexical Activation Hypothesis for Experiment 4.2 are considerably subtler than in Experiment 4.1. In particular, it is that BAs should exhibit exaggerated frequency effects compared with healthy control subjects and W/CAs should exhibit diminished (or perhaps even undetectable) frequency effects. According to Simulation Study 4.2, these effects would be realized as essentially horizontal shifts (along a VOT continuum) of the entire distribution of top-down context effects.

However, both patient groups may also suffer from bottom-up processing deficits, as seen in Experiment 4.1. Moreover, it is not altogether clear whether patients would be expected to differ from healthy controls in the extent to which they weight contextual cues (ω) during spoken word recognition. Baum (2001) showed that patients divided based on fluency (non-fluent vs. fluent) showed top-down effects of semantic sentential context in their identification responses of the first segments of words from two VOT continua (*bath–path* and *dent–tent*). While fluency does tend to correlate with clinical diagnosis (BAs tend to be non-fluent while W/CAs tend to be fluent), there were many other types of patients included in Baum's (2001) study that did not fall into the groups in question. Furthermore, the nature of the contextual cue in the present study (*the* vs. *to*, which are cues to the grammatical class of the subsequent word) is quite different from the semantic cues examined by Baum (2001). Finally, the analysis techniques employed by Baum (2001) suffer from the same issues discussed in **Section 4.3.2.3.1**; specifically, the size of boundary shifts can be influenced by both bottom-up and top-down factors, so

it remains unclear whether patients might differ from healthy control subjects in their weighting of sentential/contextual cues to the syntactic category of the target word when identifying that target word.

Whether or not patients differ from healthy controls in the size of their top-down sentential context effects, Simulation Study 4.2 illustrated distinct signatures of bottom-up, lexical-level, and sentential processing deficits. Thus, following the same approach as was taken to the analysis of Experiment 4.1, it should be possible to tease apart the impacts of different functional linguistic impairments and infer the nature of the underlying deficits in patients with BA and W/CA. To that end, Experiment 4.2 investigated top-down effects from sentential context in healthy controls and patients with BA and W/CA.

4.4.3.1. Methods

4.4.3.1.1. Subjects

Data analyzed in Experiment 4.2 came from a total of fifty subjects, 14 of whom participated in the present study as described here (8 age-matched control subjects, 3 patients diagnosed with Broca's aphasia, and 3 patients diagnosed with either Wernicke's or Conduction aphasia). The remaining data came from 36 young, healthy control subjects who participated in Experiment 1.1, described in Chapter 1 (Fox & Blumstein, in press).

Eight right-handed elderly adults (3 male) with a mean age of 73.3 years (minimum: 66; maximum: 78; sd: 4.6) participated, serving as the age-matched control (AMC) sample. All reported having age-appropriate hearing and being native speakers of English.

Three patients with Broca's aphasia (mean age: 70.3 years; minimum: 63.3; maximum: 78.1; sd: 7.4) participated, comprising the BA sample, and three patients with Wernicke's or Conduction aphasia (mean age: 69.8 years; minimum: 65.1; maximum: 78.5; sd: 7.6) participated, comprising the W/CA sample. As in Experiment 4.1, all patients' clinical diagnoses were determined based on clinical and neurological examinations (including CT scans and, where possible, MRIs) and performance on the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983). Clinical and lesion information about the patients with aphasia who participated in Experiment 4.2 is summarized in Appendix D.

The young control (YC) sample included data from 36 native monolingual speakers of American English with self-reported normal hearing who participated in Chapter 1's Experiment 1.1 (Fox & Blumstein, in press). As described in Chapter 1, a total of 50 subjects participated in the experiment. One was excluded due to technical difficulties and, of the 49 remaining subjects, 36 perceived some tokens of the *bay–pay* continuum as ambiguous (defined as making at least 10% /*b*/-responses and 10% /*p*/-responses to the two intermediate VOT tokens of the *bay–pay* continuum).

4.4.3.1.2. Stimuli

The stimuli for Experiment 4.2 were comprised of 4 acoustic tokens from a voice-onset time continuum between *bay* and *pay*, each of which was appended to a set of noun- and verb-biasing sentence contexts (e.g., *Valerie hated the...* vs. *Brett hated to...*). Beginning with a naturally produced token of *bay*, Fox and Blumstein (in press; see also Chapter 1) created a 12-step *bay–pay* VOT continuum by successively removing pitch periods from the vowel of the *bay* token and adding aspiration from a naturally-produced

pay token of equal duration between *bay*'s burst (which was amplified 2x in all tokens) and the onset of glottal pulsing. Fox and Blumstein selected 4 of the 12 tokens for inclusion in their study, corresponding to one *bay* endpoint token (VOT = 7 ms), one *pay* endpoint token (VOT = 35 ms), and two phonetically ambiguous tokens with intermediate VOTs (VOTs = 18 and 24 ms).

Minimally paired sentence contexts were selected such that, by changing only the function word that immediately preceded the target token, the contexts would create a bias for a noun (e.g., *bay*) vs. a verb (e.g., *pay*). For instance, the verb *hated* could be followed by either a noun phrase (e.g., *the bay*) or an infinitive phrase (e.g., *to pay*), so *Valerie hated the...* and *Brett hated to...* served as noun- and verb-biasing contexts, respectively. Experiment 4.2 employed a subset of 10 of Fox and Blumstein's 20 main verbs (e.g., *hate*, *want*). That reduced stimulus list can be found in Appendix E. Each of the four tokens was appended to each sentence context ending in *to...* and *the...* for a total of 80 trials (4 tokens × 10 main verbs × 2 contexts).

Note that, as mentioned earlier, data for the YCs came from Experiment 1.1 (Fox & Blumstein, in press), which included 20 main verbs (instead of 10). The full stimulus list for Experiment 1.1 can be found in Appendix A. Furthermore, although those subjects also responded to stimuli from a similarly constructed *buy–pie* continuum, the data reanalyzed here from those subjects only included their responses to the *bay–pay* continuum since it was this continuum whose VOT tokens were also stimuli in the present experiment, Experiment 4.2.

4.4.3.1.3. Procedure

All sentences were presented to participants binaurally over headphones in a random order with a 4-second inter-stimulus interval between trials. AMCs and patients with aphasia completed a minimum of 6 practice trials (some patients received more practice trials in order to adjust the volume to an appropriate level and to ensure they understood the task). Subjects were instructed to identify the last word of each sentence as either *bay* or *pay* by pressing the appropriately labeled button with their preferred hand (response mapping counterbalanced between subjects) as quickly and accurately as possible. Participants were warned that some sentences might not make sense, and they were instructed to guess if they did not know. Note that participants also completed other tasks during the same experimental session (sometimes before this task; sometimes after it).

Again, note that, since data for the YCs came from Experiment 1.1 (Fox & Blumstein, in press), the procedure differed slightly from what is described above. Most notably, YCs in Experiment 1.1 were instructed to identify the first sound of the last word in each sentence (either "b" or "p") instead of the last word of each sentence.

4.4.3.1.4. Methodological Differences Between Subject Groups

To briefly summarize the methodological disparities between the data from the YCs and the data from the elderly subjects (AMCs, BAs, and W/CAs), two are most notable: (1) the elderly subjects who participated in Experiment 4.2 heard a subset of the stimuli heard by YCs in Experiment 1.1 (and the duration of Experiment 4.2 was thus shorter) and, (2) elderly subjects performed a word identification task while YCs performed phoneme identification task (at least explicitly; see Fox & Blumstein, in press). These differences recommend caution in drawing any conclusions based on direct

comparisons of the younger and older subjects' responses. Importantly, though, the critical contrasts of interest involve comparing results among the age-matched groups of participants in Experiment 4.2 (AMCs, BAs, and W/CAs). The rationale behind the inclusion of data from Experiment 1.1 (discussed in more detail later) was that, within the hierarchical Bayesian data analysis framework, it is possible to leverage assumed commonalities between the cognitive processing underlying the two datasets, while still accounting for key differences between them.

4.4.3.2. Results: Statistical Analyses

The results of Experiment 4.2, including the results of reanalysis of data originally reported by Fox and Blumstein (in press) are shown in Figure 4.18. Recall that, to the extent that subjects tend to label stimuli with the same word-initial VOT as *pay* more often in the verb-biasing context (*to...*) than after the noun-biasing context (*the...*), those results would represent evidence of top-down effects from sentential context on speech recognition.

All statistical analyses of the present data followed exactly the approach to the logistic regression analyses taken in Experiment 4.1 (see **Section 4.3.2.2.1**). Because the two conditions being compared in the design of Experiment 4.2 differed as a function of which function word context preceded the target word, all analyses reported here included independent fixed effects for CONTEXT (β_2) (*the...* vs. *to...*; or, equivalently, noun vs. verb-biased or *bay*- vs. *pay*-biased) and for VOT (β_1) (modeled here as a continuous, linear fixed effect). No CONTEXT × VOT interaction term was included, reflecting the principle that the prior and the likelihood are independent sources of information in the Bayesian framework (*cf.* Chapter 2). Any significant main effect of

CONTEXT suggests an influence of sentential context. A significant main effect of VOT suggests that subjects' likelihood of making a *pay*-response depends on the VOT of the stimulus. Thus, these two fixed effects reflect top-down and bottom-up processing, respectively.

As in the analyses of Experiment 4.1, whenever analyses included subjects from more than one group, a fixed effect of GROUP was included in the model, along with its interactions with both CONTEXT and VOT. A significant interaction between CONTEXT and GROUP would reflect reliable differences in top-down effects of contextual information between the two groups being compared, while a significant interaction between VOT and GROUP would reflect reliable differences in bottom-up processing between the two groups. The results of Simulation Study 4.2 suggest that the weighting of lexical-level (frequency) information may be associated with the location of the inferred category boundary. Consequently, it would be difficult to determine whether differences between two groups in their best-fitting intercept coefficient (β_0) is more likely driven by differences in lexical-level processing or in the locus of their phonetic category boundaries. Moreover, interpretation of any comparisons of elderly subjects with YCs is complicated by the methodological disparities discussed in **Section 4.4.3.1.4**.

As with the analyses of Experiment 4.1, random by-subject intercepts allowed for subject variability with respect to their category boundaries (*cf.* Chapter 3). Coding of fixed effects was identical to analysis of Experiment 4.1, with a deviation-coded CONTEXT factor (contrasts: -0.5/0.5 for *the...* and *to...*, respectively) replacing the RIME factor. All results are reported in tables that include the best-fitting estimate of each

regression coefficient (β), the estimate's standard error (SE), Wald's z statistic for the estimate of that parameter ($|z|$), and the significance level of the statistic (p).

Note that the theoretical interpretations of the logistic regression coefficients are somewhat different for Experiment 4.2 than for Experiment 4.1 (see Table 4.2): β_2 reflects the influence of context, not of lexical status/frequency, and β_0 reflects not only the category boundary (a feature of the likelihood function), but also the influence of lexical-level processing (a top-down effect related to the model's prior).

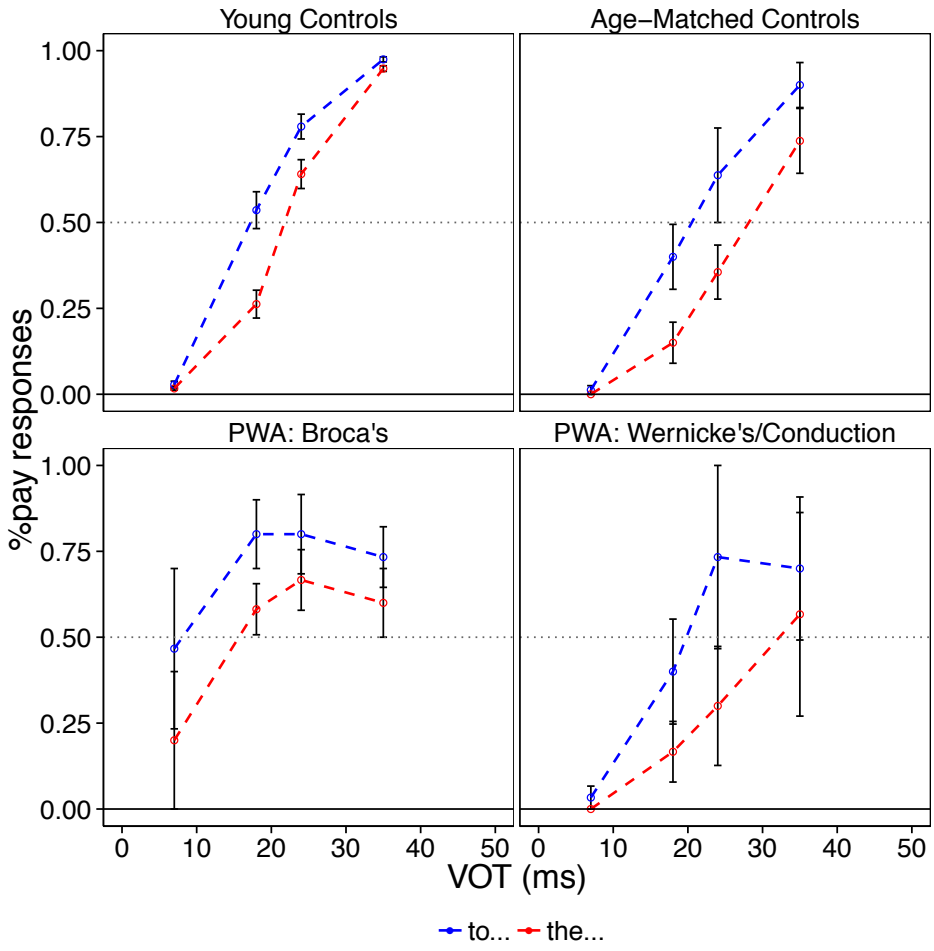

Figure 4.18. Results of Experiment 4.2: for each group, the proportion *pay*-responses as a function of VOT in the *to...* (*pay*-biased) and *the...* (*bay*-biased) conditions. Error bars represent by-subject standard error. Results for Young Controls represent reanalysis of raw data from Experiment 1.1 (Fox & Blumstein, in press). PWA = Patients with aphasia.

4.4.3.2.1. Control Subjects: YCs vs. AMCs

First, the data from all young control subjects (YCs) and all age-matched control subjects (AMCs) were submitted to logistic regression. Unsurprisingly, results (see Table 4.11) revealed significant effects from both bottom-up and top-down information sources on speech perception. Groups did not differ in the strength of their sentential context effects (β_2). The analysis also indicated that the two control groups differed in their intercept (β_0) and in the size of the effect of VOT (β_1) with results suggesting that the responses of AMCs were less influenced by VOT than YCs and that their category boundary occurred at a higher VOT value.

The former finding could be interpreted as evidence for bottom-up processing deficits (*cf.* Abada, Baum & Titone, 2008) and the latter could be interpreted as evidence of either a different phonetic category structure in older adults or as evidence for a weaker effect of lexical-level processing. However, as highlighted in **Section 4.4.3.1.4**, it is difficult to isolate the source of these disparities because several differences between these data are confounded, including at least the following: (1) the age of the participants, (2) differences in the other stimuli subjects heard during the experiment, (3) the duration of the experiment, and (4) the experimental task employed.

Coefficient	β	SE	$\lvert z \rvert$	p
β_0	-0.318	0.234	-1.359	0.174
β_1	0.260	0.010	26.492	< 0.001
β_2	1.317	0.132	9.995	< 0.001
β_0: YC vs. AMC	-1.097	0.468	-2.341	0.019
β_1: YC vs. AMC	-0.071	0.019	-3.730	< 0.001
β_2: YC vs. AMC	0.377	0.263	1.432	0.152

Table 4.11. Results of logistic regression analysis of Experiment 4.2 that included Young and Age-Matched Controls. Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary and lexical-level processing); β_1 = VOT (related to gain/slope of sigmoid); β_2 = CONTEXT (related to size of the boundary shift introduced by the contextual bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic. Results for Young Controls represent reanalysis of raw data from Experiment 1.1 (Fox & Blumstein, in press).

Given the between-group differences, follow-up tests were conducted to examine each group's data separately. Results showed that both YCs (Table 4.12) and AMCs (Table 4.13) exhibited strong effects of both bottom-up and top-down influences on speech recognition. Whatever the source of the differences between the YCs and AMCs, the results suggest that the two groups' data should not be fit together in the model-based analyses (**Section 4.4.3.3**). Based on these results, the model-based analyses fit unique values of μ_B (related to the phonetic category boundary, χ) for YCs and AMCs and allowed AMCs to have a greater category variance than YCs.

Coefficient	β	SE	$\lvert z \rvert$	p
β_0	0.230	0.197	1.167	0.243
β_1	0.297	0.008	36.418	< 0.001
β_2	1.129	0.088	12.850	< 0.001

Table 4.12. Results of logistic regression analysis of Experiment 4.2 that included only Young Controls. Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary and lexical-level processing); β_1 = VOT (related to gain/slope of sigmoid); β_2 = CONTEXT (related to size of the boundary shift introduced by the contextual bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic. Results for Young Controls represent reanalysis of raw data from Experiment 1.1 (Fox & Blumstein, in press).

Coefficient	β	SE	$\|z\|$	p
β_0	-0.865	0.415	-2.085	0.037
β_1	0.224	0.018	12.344	< 0.001
β_2	1.503	0.249	6.036	< 0.001

Table 4.13. Results of logistic regression analysis of Experiment 4.2 that included only Age-Matched Controls. Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary and lexical-level processing); β_1 = VOT (related to gain/slope of sigmoid); β_2 = CONTEXT (related to size of the boundary shift introduced by the contextual bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

4.4.3.2.2. Elderly Subjects: AMCs vs. BAs vs. W/CAs

A logistic regression examined all of the elderly participants, including the AMCs, BAs and W/CAs. Notably, this between-group comparison does not suffer from the same methodological disparities as the comparison of the YCs and AMCs. The results of this analysis are shown in Table 4.14. Overall, there was a significant top-down contextual biasing effect on subjects' responses (more *pay*-responses after verb-biasing contexts than noun-biasing contexts), and an overall effect of VOT on subjects' responses (more *pay*-responses to stimuli with longer VOTs).

Coefficient	β	SE	$\|z\|$	p
β_0	-0.531	0.364	-1.457	0.145
β_1	0.154	0.012	13.085	< 0.001
β_2	1.382	0.187	7.371	< 0.001
β_0: AMC vs. BA	2.054	1.085	1.893	0.058
β_0: AMC vs. W/CA	-1.379	1.103	-1.250	0.211
β_1: AMC vs. BA	-0.197	0.029	-6.832	< 0.001
β_1: AMC vs. W/CA	0.055	0.038	1.466	0.143
β_2: AMC vs. BA	-0.999	0.500	-1.999	0.046
β_2: AMC vs. W/CA	0.747	0.604	1.236	0.216

Table 4.14. Results of logistic regression analysis of Experiment 4.2 that included Age-Matched Controls (AMC), patients with Broca's aphasia (BA), and patients with Wernicke's or Conduction aphasia (W/CA). Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary and lexical-level processing); β_1 = VOT (related to gain/slope of sigmoid); β_2 = CONTEXT (related to size of the boundary shift introduced by the contextual bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

BAs differed from AMCs with respect to both of these effects, and there was a marginal difference ($p = 0.058$) between BAs and AMCs in their intercept. In particular, the influence of VOT on speech recognition was diminished in BAs compared to AMCs, corresponding to a shallower slope of the sigmoidal categorization curve, suggesting bottom-up processing deficits. The effect of sentential context was weaker in BAs than AMCs, suggesting impairments in the integration of sentential cues during word recognition. The marginal difference in the intercept between BAs and AMCs suggested that BAs had an inferred category boundary at a much lower VOT value than AMCs, which could correspond either to a disruption in BAs' internal phonetic category structure or increased weighting of lexical-level information. Notably, the latter interpretation is consistent with the Lexical Activation Hypothesis. Previous research has suggested fundamental aspects of phonetic category structure are preserved in aphasia (Blumstein et al, 1984; Blumstein et al, 1977b; Caplan et al, 1994; Gow & Caplan, 1996), even while discrimination, categorization, and acoustic-phonetic processing is impaired, so there is little reason to suspect that phoneme category boundaries differ between AMCs and BAs.

No significant differences were found between W/CAs and AMCs. However, the direction of the (non-significant) regression coefficient corresponding to the inferred category boundary was opposite that of the difference between BAs and AMCs. This is the direction predicted by the Lexical Activation Hypothesis.

To further examine the pattern of sentential context effects in the two patient groups, each group's data were analyzed separately. Results confirmed that both BAs (Table 4.15) and W/CAs (Table 4.16) exhibited a robust influence of sentential context in their responses. The VOT of the stimuli also influenced speech recognition in both BAs

and W/CAs, although the raw effect size for BAs was much weaker than in both of the control groups (*cf.* Tables 4.12 and 4.13).

Coefficient	β	SE	$\lvert z \rvert$	p
β_0	0.486	0.205	2.370	0.018
β_1	0.054	0.014	3.759	< 0.001
β_2	0.863	0.283	3.046	0.002

Table 4.15. Results of logistic regression analysis of Experiment 4.2 that included only patients with Broca's aphasia (BAs). Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary and lexical-level processing); β_1 = VOT (related to gain/slope of sigmoid); β_2 = CONTEXT (related to size of the boundary shift introduced by the contextual bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

Coefficient	β	SE	$\lvert z \rvert$	p
β_0	-1.292	1.111	-1.163	0.245
β_1	0.189	0.028	6.874	< 0.001
β_2	1.835	0.427	4.296	< 0.001

Table 4.16. Results of logistic regression analysis of Experiment 4.2 that included only patients with Wernicke's or Conduction aphasia (W/CAs). Shaded boxes indicate statistically significant effects. β_0 = intercept (related to phoneme category boundary and lexical-level processing); β_1 = VOT (related to gain/slope of sigmoid); β_2 = CONTEXT (related to size of the boundary shift introduced by the contextual bias); β: best-fitting estimate of each regression coefficient, SE: the estimate's standard error, $|z|$: Wald's z statistic for the estimate of that parameter, p: the significance level of the test statistic.

4.4.3.2.3. Summary of Results of Statistical Analyses

Figure 4.19 provides an alternate way of visualizing differences in the size of top-down effects from sentential context for each group over the entire continuum. For each subject's responses to each of the four VOT tokens, we computed the difference in the proportion of *pay*-responses in the verb-biased condition (*to...*) and the noun-biased condition (*the...*), and plotted the mean difference (i.e., effect size) for each group at each VOT. In summary, there are at least five tentative conclusions that find support in the statistical analyses presented above.

1. Sentential context influences speech categorization in all groups, including both patient groups.

2. For the healthy controls and for the W/CAs, those effects tend to arise most strongly at intermediate VOTs; unambiguous speech tokens less likely to be susceptible to contextual biases.
3. The behavior of YCs and AMCs appear to reflect differences in the bottom-up processing of the tokens from the VOT continua; this may be due to methodological differences in the way those two datasets were collected.
4. There is fairly robust evidence for differences between BAs and AMCs in both top-down and bottom-up speech processing, but it is difficult to draw strong conclusions from the present analyses.
5. The present analyses do not provide clear evidence for differences between behavioral response patterns of AMCs and W/CAs.

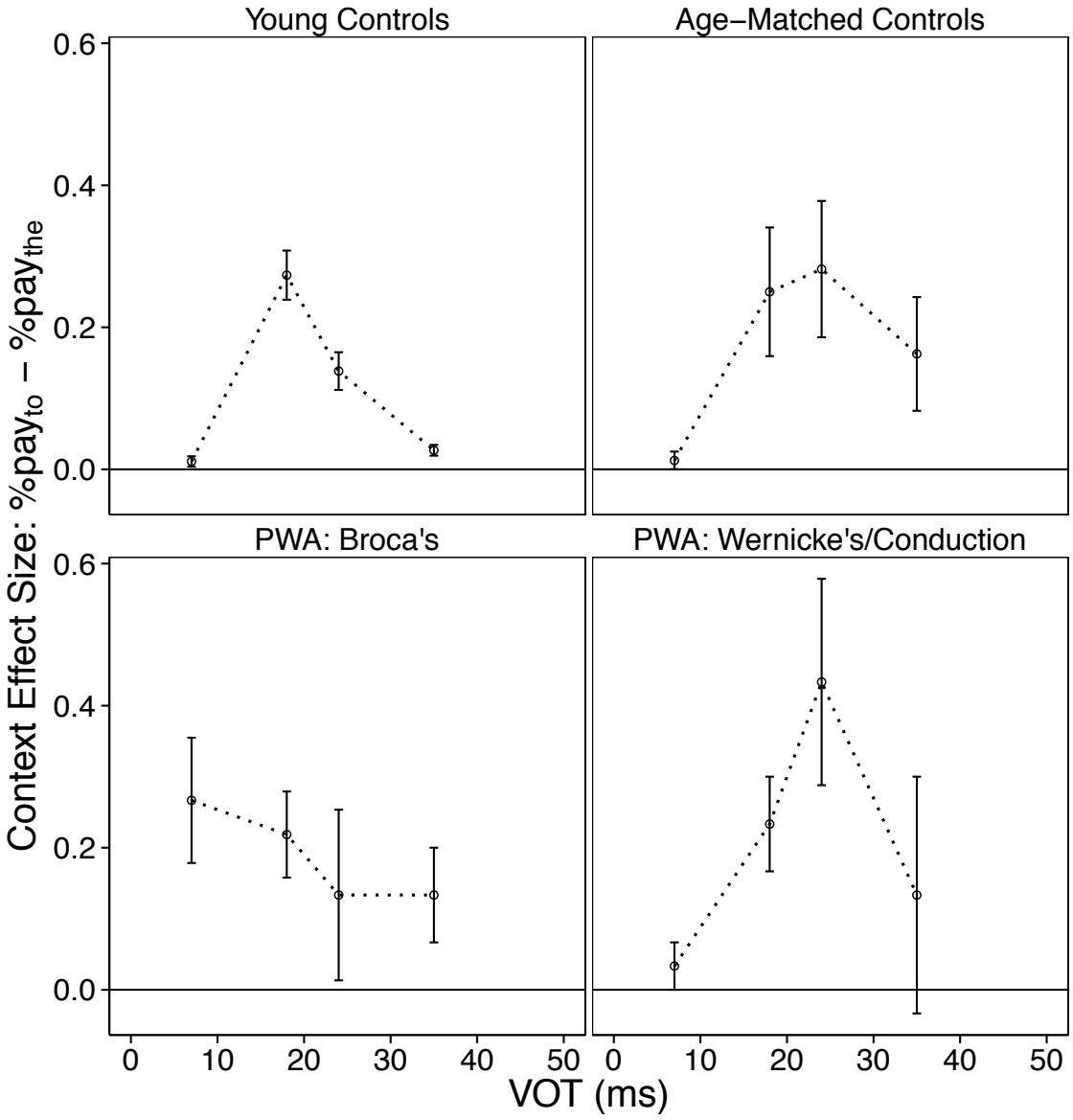

Figure 4.19. Results of Experiment 4.2: Difference between proportion *pay*-responses in the *to...* (*pay*-biased) and *the...* (*bay*-biased) conditions as a function of voice-onset time (VOT), for each group. Error bars represent by-subject standard error. Results for Young Controls represent reanalysis of raw data from Experiment 1.1 (Fox & Blumstein, in press). PWA = Patients with aphasia.

4.4.3.3. Results: Model-Based Analyses

4.4.3.3.1. Motivation of Model-Based Analyses

The statistical analyses proved difficult to interpret for several reasons. Firstly, the data for the YCs came from a different experiment than the data for the AMCs, BAs and

W/CAs. Secondly, there was limited data, with only three patients in each clinical group. Thirdly, simulations with BIASES-A in Simulation Study 4.2 suggested that the intercept parameter of the logistic regression can be influenced by both bottom-up and top-down processing components, so differences between groups were ambiguous. Moreover, the logistic regression analyses in Experiment 4.2 suffer from the same shortcomings described earlier (see **Section 4.3.2.3.1**).

Many of these problems are addressed by the model-based Bayesian data analysis approach described in **Section 4.3.2.3.1**. Its ability to distinguish between the subtle influences of many parameters while avoiding typical assumptions of many standard statistical tests (e.g., frequentist logistic regression) and its ability to make the most out of limited data by employing theoretically informed hierarchical modeling are especially advantageous for the present analyses.

4.4.3.3.2. Key Results of Model-Based Analyses

Table 4.17 provides a summary of the posterior distributions of the parameters that were fit in the present analysis (i.e., the "best-fitting" model parameters).

	Mean	SD	95% HDI min	95% HDI max
α	0.82	0.09	0.65	1.01
σ^2	221.81	5.99	210.60	233.07
YCs: μ_B	-5.95	0.18	-6.29	-5.60
Elderly: μ_B	-0.95	0.65	-2.13	0.46
AMCs: σ^2_N	85.41	24.48	39.85	135.76
BAs: σ^2_N	698.09	357.85	99.27	1467.97
W/CAs: σ^2_N	201.58	74.12	64.17	345.73
BAs: ε	0.13	0.09	9.00e-05	0.31
W/CAs: ε	0.03	0.03	5.36e-06	0.09
BAs: ϕ	2.03	0.47	1.15	3.01
W/CAs: ϕ	-0.09	0.50	-0.99	0.97
BAs: ω	0.91	0.32	0.35	1.56
W/CAs: ω	1.29	0.37	0.54	2.02

Table 4.17. Summary statistics of posterior distributions of Bayesian data analysis of Experiment 4.2. HDI = highest density interval.

The most theoretically important results regard the posterior distributions of the BAs and W/CAs and the extent to which the present model-based analysis could confidently infer differences in the posterior estimates between the patient groups and the AMCs. Three key results emerged. First, BAs were substantially more impaired in their bottom-up acoustic-phonetic processing of speech tokens (σ_N^2) compared to AMCs. Interestingly, no such difference between AMCs and W/CAs was found.

The other two key results regard each patients' weighting of lexical-level (frequency) information. Recall that optimal weighting of frequency information, which is assumed for AMCs, is given by $\phi = 1$. To the extent that ϕ can be confidently assessed to be greater than 1 for some group, it suggests that those subjects are overweighting frequency information, which is the prediction the Lexical Activation Hypothesis makes for BAs (see Table 4.1). To the extent that ϕ can be confidently assessed to be less than 1 for some group, it suggests that those subjects are underweighting frequency information, which is the prediction the Lexical Activation

Hypothesis makes for W/CAs (see Table 4.1). According to Kruschke (2011) a parameter can be confidently assessed to be different from some value if the 95% HDI of that parameter's posterior distribution excludes that value.

The second key finding was that patients with BA reliably overweight frequency information, and the third key finding was that patients with W/CA reliably underweight frequency information. Both results are exactly what is predicted by the Lexical Activation Hypothesis. This result can be seen visually in Figures 4.18 and 4.19 as the apparent shift of the entire distribution of top-down effects to the left for the BAs (to be centered over lower VOT values, which correspond to *bay*, the less frequent candidate word) or to the right for W/CAs (to be centered closer to the overall phonetic category boundary). However, among the many other differences between the behavioral patterns, it is virtually impossible to confidently asses the status of each effect's reliability without a theoretically and analytically powerful technique like the one presented here. The power of this analysis technique is that it can separate out all of the other differences between the distributions of responses and isolate the influence of each parameter on subjects' behavior.

As in the analyses of the results from Experiment 4.1, we performed a posterior predictive check (PPC) in order to evaluate the ability of the fit model to accurately capture the key aspects of the behavioral data. The PPC proceeded exactly as described in Experiment 4.1 (see **Section 4.3.2.3.2**): 100 random samples were selected from the joint posterior distribution of the model, and parameter values for a given sample were set to the sampled value in each corresponding Markov chain. For each sample we simulated data from the model, and we ran all of the statistical analyses reported in **Section**

4.4.3.3.1 on the simulated data. This yielded 100 samples of each of 6 statistical analyses. For each logistic regression coefficient in each statistical test, we computed the mean coefficient estimate ($\boldsymbol{\beta}$) and we determined how many of the statistical tests reached significance at the 0.05 level. To the extent that statistical tests on new, generated data give similar inferences as the same statistical tests on the original data, it would suggest that the model from which the data were generated captures some fundamental aspects of the generative model underlying the psychological processes giving way to the relevant empirical data.

The results are shown in Table 4.18. Figures 4.20 and 4.21 superimpose the results of the PPC onto the original experimental data shown in Figures 4.18 and 4.19. The PPCs' coefficient estimates and pattern of significances were somewhat consistent with the statistics of the original experimental data, but future work should examine possible shortcomings. The inconsistencies in this method may be related to the fact that the HDIs of some parameters were quite large (see Table 4.17), which is indicative of a dataset with inconsistent or too little data. It is also possible that inconsistencies may be due to incorrect assumptions in the model. These are important questions for future work.

Experiment 4.2: Sentential Context Effect		Results: Experiment 4.2		Results: PPC	
logistic regression	**coefficient**	***β***	***p***	**mean *β***	**% sims *p<.05***
Control Subjects	β_0	-0.318	0.174	-0.24	94
	β_1	0.260	< 0.001	0.21	100
	β_2	1.317	< 0.001	0.97	100
	β_0: YC vs. AMC	-1.097	0.019	-0.87	100
	β_1: YC vs. AMC	-0.071	< 0.001	-0.07	85
	β_2: YC vs. AMC	0.377	0.152	0.07	7
Elderly Subjects	β_0	-0.531	0.145	-0.34	92
	β_1	0.154	< 0.001	0.11	100
	β_2	1.382	< 0.001	1.01	100
	β_0: AMC vs. BA	2.054	0.058	1.64	100
	β_0: AMC vs. W/CA	-1.379	0.211	-0.98	92
	β_1: AMC vs. BA	-0.197	< 0.001	-0.14	100
	β_1: AMC vs. W/CA	0.055	0.143	0.01	12
	β_2: AMC vs. BA	-0.999	0.046	-0.41	28
	β_2: AMC vs. W/CA	0.747	0.216	0.42	15
YCs	β_0	0.230	0.243	0.20	98
	β_1	0.297	< 0.001	0.25	100
	β_2	1.129	< 0.001	0.93	100
AMCs	β_0	-0.865	0.037	-0.67	100
	β_1	0.224	< 0.001	0.18	100
	β_2	1.503	< 0.001	1.01	97
BAs	β_0	0.486	0.018	0.48	80
	β_1	0.054	< 0.001	0.04	64
	β_2	0.863	0.002	0.81	60
W/CAs	β_0	-1.292	0.245	-0.83	96
	β_1	0.189	< 0.001	0.12	100
	β_2	1.835	< 0.001	1.23	85

Table 4.18. Summary of the results of a Posterior Predictive Check (PPC) examining the reliability of the model fit to data from Experiment 4.2.

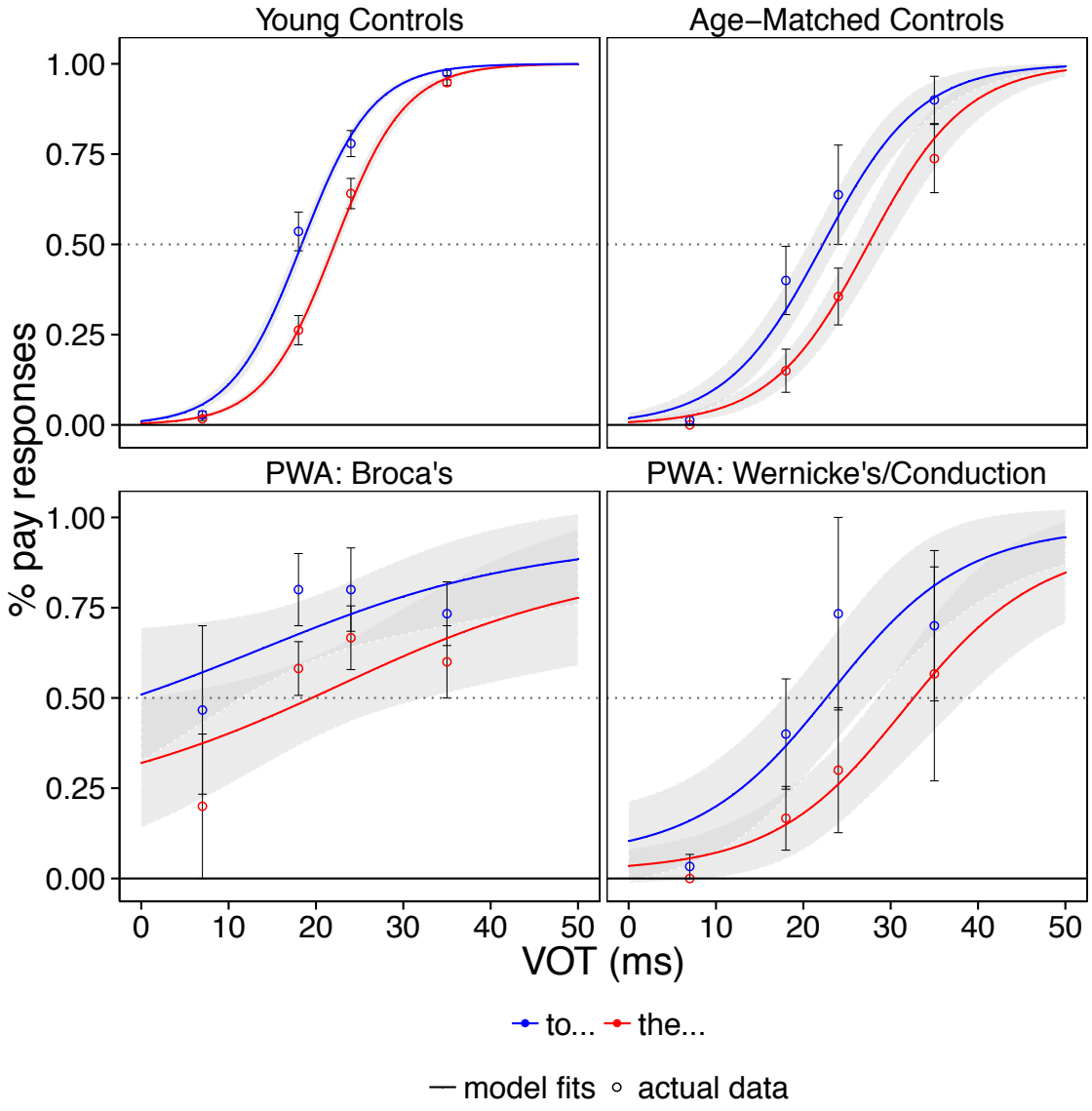

Figure 4.20. Results of Experiment 4.2 (data points; *cf.* Figure 4.18) with superimposed model fits (solid lines). For each group (panel), two curves display the two sigmoidal posterior probability functions of the *to...* (*pay*-biased) and *the...* (*bay*-biased) conditions. Points indicate proportion *pay*-responses in the *to...* (*pay*-biased) and *the...* (*bay*-biased) conditions for each VOT, for each group. Error bars represent by-subject standard error**.** PWA = Patients with aphasia.

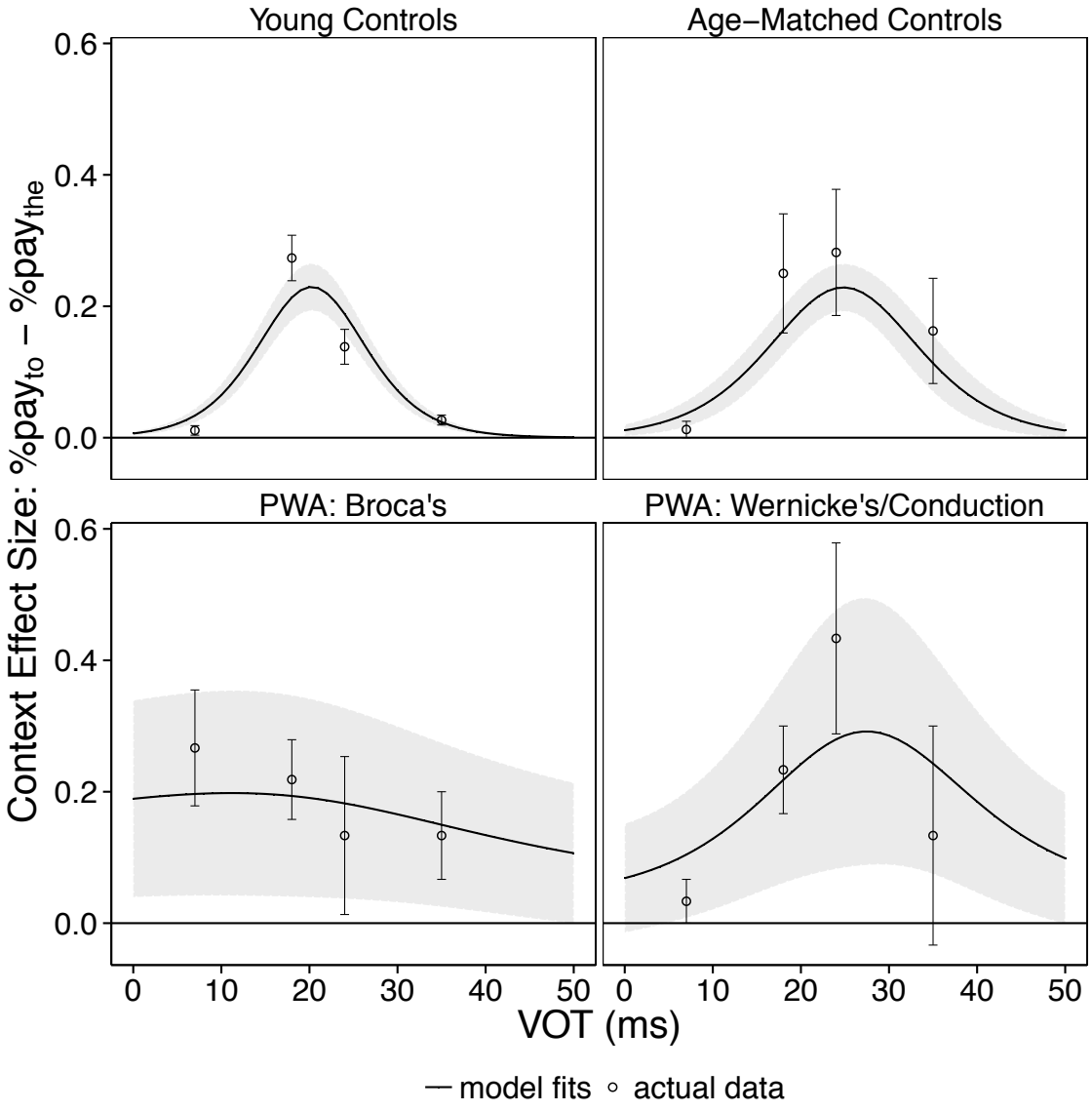

Figure 4.21. Results of Experiment 4.2 (data points; *cf*. Figure 4.19) with superimposed model fits (solid lines). For each group (panel), the curve represents the difference between proportion *pay*-responses in the *to...* (*pay*-biased) and *the...* (*bay*-biased) conditions as a function of voice-onset time (VOT). Points indicate difference in proportion *pay*-responses in the *to...* (*pay*-biased) and *the...* (*bay*-biased) conditions for each VOT, for each group. Error bars represent by-subject standard error**.** PWA = Patients with aphasia.

4.4.3.4. General Discussion of Results of Experiment 4.1 and 4.2

Together with the results of Experiment 4.1, the present work provides evidence for diminished influence of lexical status and frequency information on word recognition

in patients with W/CA, and for greater influence of lexical status and frequency information on word recognition in patients with BA. These conclusions are consistent with the original predictions of the Lexical Activation Hypothesis. At the same time, the analyses may also point towards bottom-up processing impairments (especially at the acoustic-phonetic level) in patients. However, the present results suggest that lexical processing deficits are not likely to be accounted for as downstream effects of bottom-up processing deficits alone.

Methodologically, the present results have also illustrated the relative power of hierarchical Bayesian data analysis techniques over traditional methods. Having developed a computational model, BIASES, which served as a theoretical lens through which to view the issue of lexical processing deficits in aphasia, it was possible to avoid many of the inadequate and inappropriate assumptions of more traditional statistical analyses. Ultimately, we were able to draw novel, rich, and principled conclusions from previously published data (Blumstein et al, 1994) and from another dataset with several significant limitations (e.g., it was collected in two separate experiments with different methods and it featured a relatively small number of patients and trials). This represents a promising direction for future work interested in teasing apart subtle differences in the expected influences of different model parameters on subjects' response patterns.

Finally, as discussed earlier, aphasia is a heterogeneous disorder of language. Ignoring this fact – for instance, by analyzing data without respect for their symptomology, clinical diagnosis, or underlying neurological etiology/lesion site – is not theoretically motivated and is likely to miss important differences in patients, both from each other and from healthy control subjects.

Conclusion

During auditory language comprehension, bottom-up acoustic cues in the sensory signal are critical to listeners' ability to recognize spoken words, but listeners are also sensitive to higher-level processing; in general, identification of ambiguous targets is biased by prior expectations (e.g., words over non-words, contextually consistent words over inconsistent words). The focus of the present work has been to better characterize how such top-down cues are integrated with bottom-up cues. In particular, the goal was to improve our understanding of the computational principles underlying top-down effects on speech perception, especially those top-down effects which arise from a word's sentential context.

Chapter 1 considered a longstanding debate: do top-down effects result from *interactive* modulation of perceptual processing or from entirely *autonomous*, decision-level processing? Although some past work suggested that the time course of top-down effects was incompatible with interactive models, Experiments 1.1 and 1.2 illustrated that, with appropriate controls, the predictions of interactive models were supported.

Ultimately, though, two major weaknesses of existing spoken word recognition models (whether interactive or autonomous) are that they ignore the role of sentential context and that they ignore the enormous variability in the size of top-down effects. To address these gaps, Chapter 2 introduced *BIASES* (short for *Bayesian Integration of Acoustic and Sentential Evidence in Speech*), a newly developed computational model of speech perception.

Chapter 3 demonstrated BIASES' ability to predict and explain fine-grained variability and asymmetries in previously published work, as well as in novel

experimental data from Experiment 3.1. The results of Chapter 3 indicate that many of the hallmarks of a Bayesian cue integration model are present in listeners' behavior during spoken word recognition tasks.

Finally, Chapter 4 employed BIASES to examine top-down processing in patients with aphasia. Experiment 4.1 reanalyzed previously published data (Blumstein et al, 1994) regarding top-down effects of lexical status on speech perception in patients with aphasia. Experiment 4.2 examined new data regarding top-down effects of sentential context in patients with aphasia. Model-based analysis of these data suggested that patients with aphasia experience both bottom-up processing deficits and lexical-level processing deficits, and that the lexical processing deficits are consistent with the predictions of the Lexical Activation Hypothesis (Blumstein & Milberg, 2000; McNellis & Blumstein, 2001). Importantly, those impairments differ as a function of patients' clinical diagnoses.

The BIASES model has the potential to guide future experimental research and help advance both psycholinguistic and neurolinguistic theory. This work offers new insights into the computations occurring at the interface between the perceptual processing of speech and the cognitive and linguistic processing of language.

Appendix A: Context Sentences for Experiments 1 & 2

Noun-biased (bay/pie)	Verb-biased (buy/pay)
Tom liked the...	Dennis liked to...
Jill preferred the...	Stephanie preferred to...
Valerie hated the...	Brett hated to...
Theresa chose the...	Bethany chose to...
Ronald remembered the...	Christopher remembered to...
Austin forgot the...	Rob forgot to...
Lillian neglected the...	Eliza neglected to...
Justin wanted the...	Joe wanted to...
Tina loved the...	Nathan loved to...
Noah prepared the...	Dustin prepared to...
Jasmine demanded the...	Tyler demanded to...
Josh declined the...	Grant declined to...
Celia offered the...	Kristen offered to...
Mark meant the...	Kate meant to...
Sue needed the...	Megan needed to...
Eileen expected the...	Dorothy expected to...
Katherine requested the...	Lance requested to...
Tony knew the...	Bob knew to...
Tracy promised the...	Carl promised to...
Abigail thought the...	Jacqueline thought to...

Appendix B: Filler Target Words for Experiment 2

build	put
beat	pick
break	print
blame	play
block	plan
brief	press
back	pack
bet	pet
bear	pair
bull	pull

Appendix C: Supplementary Materials

Complete Results and Discussion of Experiment 1

A.1. Details of Analysis Procedures

Because subjects' responses were categorical (/*p*/ vs. /*b*/), the data were analyzed using mixed effects logistic regression (Baayen, Davidson & Bates, 2008; Jaeger, 2008), implemented using the lme4 package (Bates, Maechler, Bolker & Walker, 2014) in *R* (R Core Team, 2014). Factorial main effects (CONTEXT, CONTINUUM, BIAS, and SPEED) were deviation-coded (contrasts: 0.5, -0.5; positive contrasts corresponded to noun-biased, *buy–pie*, /*p*/-congruent, and fast trials). VOT was a centered, continuous fixed effect. Since the design was fully within-subjects and within-items (an item corresponded to a main verb; e.g., *hated*), the maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013) for this design included all random intercepts, slopes and interactions for every subject and item. In order to achieve convergence while minimizing the risk of inferential bias (Barr et al., 2013), random correlations were excluded.

A.2. Supplementary Results/Discussion

Besides the two critical findings discussed in the main text (CONTEXT × CONTINUUM and BIAS × SPEED interactions), our results provided evidence that several other factors influence subjects' responses. Most are attributable to phonetic factors in our stimuli and well-established observations about how context effects interact with phonetic factors in speech perception.

A.2.1. Analysis 1a (omnibus): CONTEXT × CONTINUUM × VOT

In addition to the crucial CONTEXT × CONTINUUM interaction in Experiment 1, there was a main effect of VOT ($\beta = 0.42$, SE = 0.04, $|z| = 11.70$, $p < 0.001$) such that

tokens with longer VOTs were more often labeled as beginning with /*p*/, as expected given that VOT is the primary cue distinguishing the /*b*/ and /*p*/ categories in English (Liberman, Harris, Kinney & Lane, 1961).

A significant VOT × CONTEXT interaction (β = 0.10, SE = 0.03, $|z|$ = 3.58, p < 0.001) replicates previous work showing that the size of a top-down bias depends on the acoustic ambiguity of the stimuli (Burton, Baum & Blumstein, 1989; Ganong, 1980; McQueen, 1991; Pitt & Samuel, 1993; Tuinman et al, 2014; van Alphen & McQueen, 2001). As Figure 1.1 suggests, the closer a token's mean rate of /*p*/-responses was to the phoneme category boundary (the VOT at which one would expect to see 50% /*b*/-responses and 50% /*p*/-responses), the larger the difference between subjects' /*p*/-response rates at the two levels of CONTEXT appears to be.

VOT interacted with CONTINUUM (β = 0.18, SE = 0.03, $|z|$ = 6.30, p < 0.001), suggesting a somewhat stronger influence of VOT in the *buy–pie* continuum than in the *bay–pay* continuum. Although the exact source of this asymmetry is not immediately obvious, one should not necessarily expect the effect of VOT to pattern identically in the *bay–pay* and *buy–pie* continua, because VOT is only one of many cues to the identity of phonetically ambiguous (between /*b*/ and /*p*/) stimuli. Burst amplitude (Repp, 1984), subsequent vowel duration (Miller & Dexter, 1988; Summerfield, 1981), vowel identity (Klatt, 1975; Stevens & Klatt, 1974), and the lexical frequency of continuum endpoints (Fox, 1984) can all influence voicing decisions about phonetically ambiguous stimuli, and although it is not clear which of these (if any) contributed to this asymmetry in Experiment 1, it is unclear how any of these factors could account for the theoretically important CONTEXT × CONTINUUM interaction.

Finally, a significant main effect of CONTEXT (β = -0.36, SE = 0.12, $|z|$ = 2.89, p < 0.004) such that subjects were more likely to make /*p*/-responses after verb-biasing sentences than after noun-biasing sentences reflected the fact that the simple effect of CONTEXT was stronger in the *bay–pay* continuum (β = -1.37; /*p*/-responses to ambiguous tokens: 44.4% in noun-biased contexts vs. 65.5% in verb-biased contexts) than in the *buy–pie* continuum (β = 0.95; /*p*/-responses to ambiguous tokens: 58.4% in noun-biased contexts vs. 38.5% in verb-biased contexts). Further research would be necessary to identify the specific source of this asymmetry, but one possibility is that the syntactic manipulation was more efficacious in the *bay–pay* continuum because the specific items in the experiment (e.g., *Brett hated to...*) created stronger preferences when judging between *bay* and *pay* than between *buy* and *pie*. Importantly, though, no matter the cause of this or any of the other ancillary effects discussed here, the prediction that CONTEXT would have robust, contrasting effects in the two continua was borne out by the data.

A.2.2. Analysis 1b (follow-up tests): CONTEXT × VOT

In addition to the reported simple effects of CONTEXT in the by-continuum follow-up tests, both analyses, as in the omnibus analysis, revealed simple effects of VOT (*bay–pay*: β = 0.33, SE = 0.03, $|z|$ = 11.58, p < 0.001; *buy–pie*: β = 0.50, SE = 0.04, $|z|$ = 13.74, p < 0.001) in the expected direction. Finally, there was a significant interaction between CONTEXT and VOT in the *bay–pay* continuum (β = 0.12, SE = 0.03, $|z|$ = 3.48, p < 0.001) and marginal interaction in the *buy–pie* continuum (β = 0.07, SE = 0.04, $|z|$ = 1.80, p = 0.07), suggesting that ambiguous tokens were differentially impacted by CONTEXT in both continua (see Appendix C for discussion).

A.2.3. Analysis 2: BIAS × SPEED × VOT

In addition to the BIAS × SPEED interaction, the results revealed a main effect of VOT (β = 0.36, SE = 0.03, $|z|$ = 12.47, $p < 0.001$) and a main effect of BIAS (β = 0.99, SE = 0.14, $|z|$ = 7.21, $p < 0.001$), such that /*p*/-responses were more likely when targets had longer VOTs and in trials for which the CONTEXT and CONTINUUM jointly made /*p*/ the congruent response. There was also a significant VOT × SPEED interaction (β = 0.09, SE = 0.03, $|z|$ = 3.12, $p < 0.002$), suggesting that slower responses to a token were less influenced by that token's VOT.

Complete Results and Discussion of Experiment 2

B.1. Details of Analysis Procedures

Analyses followed the same approach as Experiment 1's. When occasional convergence failures occurred, the random effects structure was simplified by removing random slopes for factors involving the VOT factor. In all cases, this simplification allowed for convergence, and the pattern of results (i.e., which fixed effects reached significance) was identical to the results of the unconverged models with all of the random effects.

B.2. Supplementary Results/Discussion

Besides the two critical findings of Experiment 2 discussed in the main text (CONTEXT × CONTINUUM interaction, but no BIAS × SPEED interaction), there was evidence that several other effects influenced subjects' responses, including effects replicating most patterns seen in Experiment 1 (see above). However, aside from differing in the presence of a BIAS × SPEED interaction, Experiments 1 and 2 differed in a few other ways. In particular, the critical tokens were, on the whole, less often identified

as /*p*/ in Experiment 2 than in Experiment 1, even for the /*p*/-endpoint tokens (*pay*-endpoint: 96.0% vs. 77.0% /*p*/-responses in Experiment 1 vs. 2; *pie*-endpoint: 86.8% vs. 76.5%). Such a pattern is consistent with earlier studies showing that the distributional statistics of acoustic-phonetic cues (e.g., VOTs) within an experimental context can produce range effects in the perception of phonetic category structure (Clayards, Tanenhaus, Aslin & Jacobs, 2008). An analysis of the VOTs of the ten naturally produced /*p*/-initial filler targets showed that these fillers had a mean VOT of 90 ms (with the shortest VOT being 71 ms), in contrast to 35 and 34 ms VOTs for the two critical /*p*/-endpoint stimuli. Thus, it appears that the longer VOTs of the filler targets affected the perception of voicing in the critical target stimuli such that the boundary between the */p/* and */b/* stimuli was now skewed towards fewer */p/* and more */b/* responses, consistent with range effects (Brady & Darwin, 1978).

B.2.1. Analysis 1a (omnibus): CONTEXT × CONTINUUM × VOT

In addition to the crucial CONTEXT × CONTINUUM interaction in Experiment 2, main effects of CONTEXT (β = -0.67, SE = 0.20, $|z|$ = 3.28, $p < 0.002$) and VOT (β = 0.22, SE = 0.05, $|z|$ = 4.91, $p < 0.001$) emerged in Experiment 2, both matching patterns observed in Experiment 1.

B.2.2. Analysis 1b (follow-up tests): CONTEXT × VOT

Follow-up tests in each continuum revealed significant simple effects of CONTEXT (see main text), as well as simple effects of VOT (*bay–pay*: β = 0.21, SE = 0.04, $|z|$ = 5.75, $p < 0.001$; *buy–pie*: β = 0.21, SE = 0.05, $|z|$ = 4.65, $p < 0.001$).

B.2.3. Analysis 2: BIAS × SPEED × VOT

Although there was no evidence for a BIAS × SPEED interaction, the results showed a main effect of BIAS (β = 1.01, SE = 0.21, $|z|$ = 4.73, $p < 0.001$) which corresponds to the CONTEXT × CONTINUUM interaction in the primary analysis, and a main effect of VOT (β = 0.18, SE = 0.03, $|z|$ = 5.40, $p < 0.001$).

Appendix D: Patient Characteristics for Experiment 4.2

Subj	Diagnosis	Sex	Handed-ness	Age at test	Time post onset	Education (years)	Etiology	BDAE Subtests: Aud. Comp.; Fluency; Artic.; Repet'n (W/H/L)	Lesion analysis
1	Broca	M	R	63.3	10 yrs	16	Left hemisphere frontal lesion including insular cortex and lateral putamen with anterior extension into approximately half of Broca's area with deep extension across the anterior limb of the internal capsule and patchy lesion into a portion of medial subcallosal fasciculus...	*z* = +1.07 52.5 40th % 9 / 8 / 2	0% WA 0% SMG 39% IFG (by analysis)
2	Broca	M	R	69.7	27 yrs	16	Left hemisphere lesion involving caudate and global pallidus, anterior internal capsule to medial temporal cortex and insula, and anterior PVWM	*z* = +0.96 55 45th% 9 / 6 / 4	none available
3	Broca	M	L	78.1	18 yrs	16	Left hemisphere lesion including all of Broca's area and the white matter deep into it with...	*z* = +0.82 47.5 70th % 9 / 6 / 5	100% IFG (by eye)
4	Wernicke → Conduction	F	R	78.5	5 yrs	12	Wernicke's area...no SMG or Broca's area	*z* = +0.85 62.5 70th % 9 / 7 / 3	5% AG 3% WA 5% other temp. 0% SMG 0% IFG (by analysis)
5	Conduction	M	R	62*	10 yrs	16	L CVA: Posterior temporoparietal lesion involves the insula and 1/4 of Wernicke's area with superior extension into the supramarginal and angular gyrus areas	*z* = +0.52 64 60th % 7 / 1 / 1	74% WA 99% SMG
6	Wernicke	M	R	65.7	19 yrs	18	Left parietal AVM clipped; subarachnoid hemmorage...patchy Wernicke's area; extension into posterior SMG and angular gyri	*z* = +0.54 72.5 70th % 9 / 6 / 6	35% WA 18% SMG 0% IFG (by eye)

Table A.1. All parameters from model-based analyses of Experiments 1 and 2

Appendix E: Sentence Contexts for Experiment 4.2

Noun-biased (*bay*)	Verb-biased (*pay*)
Theresa chose the...	Bethany chose to...
Jasmine demanded the...	Tyler demanded to...
Valerie hated the...	Brett hated to...
Tom liked the...	Dennis liked to...
Sue needed the...	Megan needed to...
Celia offered the...	Kristen offered to...
Jill preferred the...	Stephanie preferred to...
Ronald remembered the...	Christopher remembered to...
Katherine requested the...	Lance requested to...
Justin wanted the...	Joe wanted to...

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological review*,*116*(3), 463.

Andruski, J., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*, 163-187.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163-187.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Baker, E., Blumstein, S. E., & Goodglass, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia*, *19*(1), 1-15.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255-278.

Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex*, *13*(1), 85-95.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. <URL: http://CRAN.R-project.org/package=lme4>.

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, *10*(4), 417-429.

Bicknell, K., Jaeger, T. F., & Tanenhaus, M. K. (2015, in press). Now or...later: Perceptual data is not immediately forgotten during language processing. *Behavioral and Brain Sciences, 38*.

Bicknell, K., Tanenhaus, M. K., & Jaeger, T. F. (2015). Listeners can maintain and rationally update uncertainty about prior words. Manuscript submitted for publication.

Blumstein, S. E., & Milberg, W. P. (2000). Language deficits in Broca's and Wernicke's aphasia: A singular impairment. *Language and the brain: Representation and processing*, 167-184.

Blumstein, S. E. (2007). Word recognition in aphasia. In G. Gaskell (Ed.), Oxford Handbook of Psycholinguistics.

Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, *15*(1), 19-30.

Blumstein, S. E., Burton, M., Baum, S., Waldstein, R., & Katz, D. (1994). The role of lexical status on phonetic categorization in apashia. *Brain and Language, 46*, 181-197.

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: an fMRI investigation of phonetic category structure. *Cognitive Neuroscience, Journal of*, *17*(9), 1353-1366.

Brady, S. A. & Darwin, C. J. (1978). Range effects in the perception of voicing. *Journal of the Acoustical Society of America, 63*(5), 1556-1558.

Burton, M. W., Baum, S., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 567-575.

Burton, M. W., & Blumstein, S. E. (1995). Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1230-1235.

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Cognitive Neuroscience, Journal of*, *12*(4), 679-690.

Caplan, D., & Utman, J. A. (1994). Selective acoustic phonetic impairment and lexical access in an aphasic patient. *The Journal of the Acoustical Society of America*, *95*(1), 512-517.

Carpenter, R. L., & Rutherford, D. R. (1973). Acoustic cue discrimination in adult aphasia. *Journal of Speech, Language, and Hearing Research*, *16*(3), 534-544.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive psychology*, *38*(2), 191-258.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in cognitive sciences*, *10*(7), 287-291.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues.*Cognition*, *108*(3), 804-809.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, *108*(1), 204-256.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. Journal of Memory and Language, 26 (5), 527-538.

Connine, C. M. (1990). Recognition of spoken words with ambiguous word initial phonemes. *Bulletin of the Psychonomic Society*, *28*, 497.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constrainst. *Journal of Memory and Language*, *30*(2), 234-250.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition?. *Journal of Memory and Language*, *32*(2), 193-210.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics*, *56*(6), 624-636.

Connine, C. M. & Clifton, C. C., Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 291-299.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 81.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*(3), 371-414.

Csépe, V., Osman-Sági, J., Molnár, M., & Gósy, M. (2001). Impaired speech perception in aphasic patients: event-related potential and neuropsychological assessment. *Neuropsychologia*, *39*(11), 1194-1208.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*(2), 141-177.

Cutler, A. & Norris, D. (1979). Monitoring sentence comprehension. In: *Sentence processing: Psycholinguistic studies presented to Merrill Garrett,* ed. W. E. Cooper & E. C. T. Walker. Erlbaum.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668-3678.

Dagan, I., Marcus, S., & Markovitch, S. (1993, June). Contextual word similarity and estimation from sparse data. In *Proceedings of the 31st annual meeting on Association for Computational Linguistics* (pp. 164-171). Association for Computational Linguistics.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements.*Cognitive psychology*, *42*(4), 317-367.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507-534.

De Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, *4*(4), 129-134.

De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213-231.

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and cognitive processes*, *23*(4), 583-608.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception.*Ecological Psychology*, *1*(2), 121-144.

Do, Y. A. (2011). Interaction of the top-most and the bottom-most: Pragmatic bias and phonetic perception. Presented at the *37th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, CA, Feb. 12-13.

Docherty, G. J., Watt, D., Llamas, C., Hall, D., & Nycz, J. (2011). Variation in voice onset time along the Scottish-English border. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 591-594).

Dumais, S. T. (2004). Latent semantic analysis. *Annual review of information science and technology*, *38*(1), 188-230.

Eggert, G. H. (1977). *Wernicke's works on aphasia: a sourcebook and review*. Mouton de Gruyter.

Elman, J. L. & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. Journal of Memory and Language, 27, 143-165.

Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. Psychological Review, 116 (4), 752-782.

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological review*,*120*(4), 751.

Fellbaum, C. (1998). *WordNet*. Blackwell Publishing Ltd.

Fine, A. B., & Florian Jaeger, T. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, *37*(3), 578-591.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension.

Foss, D. J., & Swinney, D. A. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*, *12*(3), 246-257.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of “new” and “old” words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*(5), 489-504.

Fowler, C. A., & Rosenblum, L. D. (1991). The perception of phonetic gestures.*Modularity and the motor theory of speech perception*, 33-59.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. Journal of Experimental Psychology: Human Perception and Performance, 10 (4), 526-540.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474-496.

Ganong, W. F., III (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance, 6 (1), 110-125.

Garnes, S. & Bond, Z. S. (1976). The relationship between semantic expectation and acoustic information. Phonologica, 3, 285-293.

Garvey, C. & Caramazza, A. (1974). Implicit causality in verbs. Linguistic Inquiry, 5, 459-464.

Gaskell, M. G., & Marslen–Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science*, *23*(4), 439-462.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive psychology*, *45*(2), 220-266.

Geisler, W. S., & Kersten, D. (2002). Illusions, perception and Bayes. *nature neuroscience*, *5*(6), 508-510.

Goodglass, H. (1993). Understanding aphasia: Foundations of neuropsychology. San Diego, CA: Academic Press.

Goodglass, H., Gleason, J. B., & Hyde, M. R. (1970). Some dimensions of auditory language comprehension in aphasia. *Journal of Speech, Language, and Hearing Research*, *13*(3), 595-606.

Gow Jr, D. W., & Caplan, D. (1996). An examination of impaired acoustic–phonetic processing in aphasia. *Brain and Language*, *52*(2), 386-407.

Grossberg, S. (1980). How does the brain build a cognitive code? *Psychological Review, 87*, 1-51.

Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, *31*(3), 423-445.

Grossberg, S. & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects.. Psychological Review, 107, 735-767.

Guediche, S., Reilly, M., & Blumstein, S. E. (2014). Facilitating perception of speech in babble through conceptual relationships. *The Journal of the Acoustical Society of America*, *135*(4), 2257-2258.

Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. Journal of Cognitive Neuroscience, 25 (5), 706-718.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. Proceedings of the North American Association for Computational Linguistics, , 159-166.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?. *Cognition*, *59*(1), 91-117.

Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, *22*(7), 1114-1139.

Hunnicutt, S. (1985). Intelligibility versus redundancy-conditions of dependency.*Language and Speech*, *28*(1), 47-56.

Isenberg, D., Walker, E. C. T., & Ryder, J. (1980). A top-down effect on the identification of function words. Paper presented at the annual meeting of the Acoustical Society of America, Los Angeles, CA.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434-446.

Janse, E. (2006). Lexical competition effects in aphasia: Deactivation of lexical candidates in spoken word processing. Brain and Language, 97 (1), 1-11.

Jauhiainen, T., & Nuutila, A. (1977). Auditory perception of speech and speech sounds in recent and recovered cases of aphasia. *Brain and Language*, *4*(4), 572-579.

Jelinek, F., & Mercer, R. (1985). Probability distribution estimation from sparse data. *IBM Technical Disclosure Bulletin*, *28*, 2591-2594.

Jelinek, F. (1990). Self-organized language modeling for speech recognition.*Readings in speech recognition*, 450-506.

Frederick Jelinek. (1997). *Statistical methods for speech recognition*. MIT press.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133-156.

Khaitan, P., & McClelland, J. L. (2010). Matching exact posterior probabilities in the Multinomial Interactive Activation Model. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (p. 623). Austin, TX: Cognitive Science Society.

Kim, D., Stephens, J. D. W., & Pitt, M. A. (2012). How does context play a part in splitting words apart? Production and perception of word boundaries in casual speech . Journal of Memory and Language, 66, 509-529.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research*,*18*(4), 686-706.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of phonetics*, *7*(312), 1-26.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel.*Psychological review*, *122*(2), 148.

Knill, D. C., Kersten, D., & Yuille, A. (1996). A Bayesian formulation of visual perception. In D. C. Knill and W. Richards (Eds.), Perception as Bayesian Inference, Cambridge University Press.

Korneef, A. W. & van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. Journal of Memory and Language, 54, 445-465.

Kronrod, Y., Coppess, E., & Feldman, N. H. (2012). A unified model of categorical effects in consonant and vowel perception. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 629-634).

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological Review, 104, 211-240.

Leeper, H. A., Shewan, C. M., & Booth, J. C. (1986). Altered acoustic cue discrimination in Broca's and conduction aphasics. *Journal of communication disorders*, *19*(2), 83-103.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, *74*(6), 431.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 54, 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of experimental psychology*, *61*(5), 379.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and speech*, *6*(3), 172-187.

Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. Word, 20, 384-422.

Lisker, L. (1986). ""Voicing"" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. Language and Speech, 29, 3-11.

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

Luce, R. D. (1959). Individual choice behavior. Oxford, England: John Wiley.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & psychophysics*,*62*(3), 615-625.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133-156.

Magnuson, J. S., Mirman, D., & Harris, H. D. (2012). Computational models of spoken word recognition. *The Cambridge handbook of psycholinguistics*, 76-103.

Magnuson, J. S., Mirman, D., & Myers, E. (2013). Spoken word recognition.*Oxford handbook of cognitive psychology*, 412-441.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons.*Journal of Experimental Psychology: General*, *132*(2), 202.

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics*, *19*(2), 313-330.

Marr, D. (1982). Vision: A computational Investigation into the human representation and processing of visual information. San Francisco: W. H. Freeman & Co.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1-71.

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. . Cognitive Psychology, 10, 29-63.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. Science, 189 (4198), 226-228.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. Cognition, 25, 71-102.

Martin, A. E., Monahan, P. J., & Samuel, A. (2012). Vowel identification shaped by phrasal gender agreement expectation. Poster at the 25th Annual CUNY Conference on Human Sentence Processing, New York, NY.

Massaro, D. W. & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception . The Journal of the Acoustical Society of America, 67, 996-1013.

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, *21*(3), 398-421.

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 960-977.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings.*Psychological review*, *88*(5), 375.

McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). Parallel distributed processing. *Explorations in the microstructure of cognition*, *2*.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 1–36). London, UK: Erlbaum.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, *23*(1), 1-44.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11-38.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception?. *Trends in cognitive sciences*, *10*(8), 363-369.

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition.*Cognitive science*, *38*(6), 1139-1189.

McDonald, J. L. & MacWhinney, B. (1995). The time course of anaphora resolution: Effects of implicit verb causality and gender. Journal of Memory and Language, 34, 543-566.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological science*,*14*(6), 648-652.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, *43*(16), 1735-1751.

McGurk, H. & McDonald, J. (1976). Hearing lips and seeing voices. Nature, 264 (5588), 746-748.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental science*, *12*(3), 369-378.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic bulletin & review*, *15*(6), 1064-1071.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of memory and language*, *60*(1), 65-91.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. Cognition, 86, B32-B42.

McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. Journal of Psycholinguistic Research, 32, 77-97.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. Journal of

Experimental Psychology: Human Perception and Performance, 34 (6), 1609-1631.

McNellis, M. & Blumstein, S. E. (2001). Self-organizing dynamics of lexical access in normals and aphasics. Journal of Cognitive Neuroscience, 13, 151-170.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity. Journal of Experimental Psychology: Human Perception and Performance, 17, 433-443.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes?.*Journal of Memory and Language*, *61*(1), 1-18.

McQueen, J. M., Norris, D., & Cutler, A. (2006). Are there really interactive processes in speech perception? *Trends in Cognitive Science, 10(12)*, 533.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 621.

Mertus, J. (1989). BLISS User Manual. Providence, RI: Brown University.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, *90*(2), 227.

Miceli, G., Caltagirone, C., Gainotti, G., & Payer-Rigo, P. (1978). Discrimination of voice versus place contrasts in aphasia. *Brain and Language*, *6*(1), 47-51.

Miceli, G., Gainotti, G., Caltagirone, C., & Masullo, C. (1980). Some aspects of phonological impairment in aphasia. *Brain and language*, *11*(1), 159-169.

Milberg, W. & Blumstein, S. E. (1981). Lexical decision and aphasia: evidence for semantic processing. Brain and Language, 14, 371-385.

Milberg, W., Blumstein, S. E., & Dworetzky, B. (1988a). Phonological factors in lexical access: Evidence from an auditory lexical decision task. Psychonomic Bulletin and Review, 26 (4), 305-308.

Milberg, W., Blumstein, S. E., & Dworetzky, B. (1988b). Phonological processing and lexical access in aphasia. Brain and Language, 34, 279-293.

Miller, J. L. & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. Journal of Experimental Psychology: Human Perception and Performance, 14 (3), 369-378.

Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition*,*41*(1), 197-229.

Miller, J. L. & Voltalis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. Perception and Psychophysics, 46, 505-512.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database*. *International journal of lexicography*, *3*(4), 235-244.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on a word identification. Perception and Psychophysics, 36 (4), 329-337.

Mirman, D., & Britt, A. E. (2014). What we talk about when we talk about access deficits. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1634).

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: evidence from eye-tracking and computational modeling. *Brain and language*, *117*(2), 53-68.

Morton, J. (1969). Interaction of information in word recognition. *Psychological review*, *76*(2), 165.

Movellan, J. R., & McClelland, J. L. (2001). The Morton-Massaro law of information integration: Implications for models of perception. *Psychological Review*, *108*(1), 113.

Myers, E. B. & Blumstein, S. E. (2008). The neural bases of the lexical effect: an fMRI investigation. *Cerebral Cortex, 18*, 346-355.

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*.

Nearey, T. M. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America*, *101*(6), 3241-3254.

Norris, D. & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continous speech recognition. Psychological Review, 115 (2), 357-395.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. Cognition, 52, 189-234.

Norris, D. & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357-395.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. Journal of Experimental Psychology: Human Perception and Performance, 21 (5), 1209-1228.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences, 23, 299-370.

Oden, G. C. & Massaro, D. W. (1978). Integration of feature information in speech perception . Psychological Review, 85, 172-191.

Ostrand, R., Blumstein, S. E., & Morgan, J. L. (2011). When hearing lips and seeing voices becomes perceiving speech: Auditory-visual integration in lexical access. In *Proceedings of the Annual Meeting of the Cognitive Science Society*(Vol. 33, pp. 1376-1381).

Pickett, J. M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*,*6*(3), 151-164.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*(2), 328-333.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Attention, Perception, & Psychophysics*, *15*(2), 285-290.

Pitt, M. A. & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. Journal of Experimental Psychology: Human Perception and Performance, 19 (4), 699-725.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, *113*(1), 57.

Pollack, I., Rubenstein, H., & Decker, L. (1960). Analysis of incorrect responses to an unknown message set. *The Journal of the Acoustical Society of America*,*32*(4), 454-457.

Prather, P. A., Zurif, E., Love, T., & Brownell, H. (1997). Speed of lexical activation in nonfluent Broca's aphasia and fluent Wernicke's aphasia. *Brain and Language*, *59*(3), 391-411.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <URL: http://www.R-project.org/>.

Recchia, G., Sahlgren, M., Kanerva, P., & Jones, M. N. (2015). Encoding Sequential Information in Semantic Space Models: Comparing Holographic Reduced Representation and Random Permutation. *Computational intelligence and neuroscience*, *2015*.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*,*92*(1), 81.

Repp, B. H. (1983). Coarticulation in sequences of two nonhomorganic stop consonants: perceptual and acoustic evidence. *The Journal of the Acoustical Society of America*, *74*(2), 420-427.

Repp, B. H. (1984). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. *Language and speech*, *27*(3), 245-254.

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature‐based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303-345.

Robson, H., Keidel, J. L., Ralph, M. A. L., & Sage, K. (2012). Revealing and quantifying the impaired phonological analysis underpinning impaired comprehension in Wernicke's aphasia. *Neuropsychologia*, *50*(2), 276-288.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Rohde, H. & Ettlinger, M. (2012). Integration of pragmatic and phonetic cues in spoken word recognition. Journal of Experimental Psychology: Learning, Memory and Cognition, 38 (4), 967-983.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review*, *89*(1), 60.

Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology, 62*, 49-72.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology.*Journal of Experimental Psychology: General*, *110*(4), 474.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes?. *Journal of Experimental Psychology: General*,*125*(1), 28.

Savin, H. B. (1963). Word‐Frequency Effect and Errors in the Perception of Speech. *The Journal of the Acoustical Society of America*, *35*(2), 200-206.

Sawusch, J. R., & Jusczyk, P. W. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human perception and performance*, *7*(2), 408.

Scharenborg, O., Norris, D., Bosch, L., & McQueen, J. M. (2005). How should a speech recognizer work?. *Cognitive Science*, *29*(6), 867-918.

Smits, R., Warner, N., McQueen, J. M., & Cutler, A. (2003). Unfolding of phonetic information over time: A database of Dutch diphone perception. *The Journal of the Acoustical Society of America*, *113*(1), 563-574.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory.

John, M. F. S., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*(1), 217-257.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced‐voiceless distinction for stops. *The Journal of the Acoustical Society of America*,*55*(3), 653-659.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, *111*(4), 1872-1891.

Strand, J., Simenstad, A., Cooperman, A., & Rowe, J. (2014). Grammatical context constrains lexical competition in spoken word recognition. *Memory & cognition*, *42*(4), 676-687.

Summerfield, Q. (1981). Articulatory rate and perceputal constancy in phonetic perception. Journal of Experimental Psychology: Human Perception and Performance, 7 (5), 1074-1095.

Swinney, D., Prather, P., & Love, T. (2000). The time-course of lexical access and the role of context: Converging evidence from normal and aphasic processing. *Language and the brain: Representation and processing*, 273-292.

Szostak, C. M., & Pitt, M. A. (2013). The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics*, *75*(7), 1533-1546.

Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, *22*(3), 259-282.

Tanenhaus, M. K. (2007). Eye movements and spoken language processing.*Eye movements: A window on mind and brain*, 309-26.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557-580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

Thomas, M. S., & McClelland, J. L. (2008). Connectionist models of cognition.*Cambridge handbook of computational cognitive modelling*, 23-58.

Toscano, J. C. & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics . Cognitive Science, 34, 434-464.

Toscano, J. C. & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization . Attention, Perception and Psychophysics, 74, 1284-1301.

Tuinman, A., Mitterer, H., & Cutler, A. (2014). Use of syntax in perceptual compensation for phonological reduction. *Language and speech*, *57*(1), 68-85.

Utman, J. A., Blumstein, S. E., & Sullivan, K. (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. Brain and Language, 79, 444-472.

Utman, C. H. (1997). Performance effects of motivational state: A meta-analysis. *Personality and Social Psychology Review*, *1*(2), 170-182.

Utman, J. A., Blumstein, S. E., & Burton, M. (2000). Effects of subphonetic and syllable structure variation on word recognition. Perception and Psychophysics, 62, 1297-1311.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech.*Proceedings of the National Academy of Sciences*, *104*(33), 13273-13278.

van Alphen, P. & McQueen, J. M. (2001). The time-limited influence of sentential context on function word identification. Journal of Experimental Psychology: Human Perception and Performance, 27 (5), 1057-1071.

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of cognitive neuroscience*, *20*(4), 580-591.

Vitevitch, M. S. & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception.. Psychological Science, 9, 325-329.

Vitevitch, M. S. & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. Journal of Memory and Language, 40, 374-408.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms.*Language and Speech*, *40*(3), 211-228.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories.*The Journal of the Acoustical Society of America*, *92*(2), 723-735.

Warner, N., Smits, R., McQueen, J. M., & Cutler, A. (2005). Phonological and statistical effects on timing of speech perception: Insights from a database of Dutch diphone perception. *Speech Communication*, *46*(1), 53-72.

Warren, R. M., & Obusek, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, *9*(3), 358-362.

Warren, R. M. & Sherman, G. L. (1974). Phonemic restorations based on subsequent context. Perception and Psychophysics, 16, 150-156.

Warren, R. M. & Warren, R. P. (1970). Auditory illusions and confusions. Scientific American, 223, 30-36.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. Science, 167, 392-393.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. The Journal of the Acoustical Society of America, 60, 1381-1389.

Yee, E., Blumstein, S. E., & Sedivy, J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *Journal of cognitive neuroscience*, *20*(4), 592-612.

Yeni-Komshian, G. H., & Lafontaine, L. (1983). Discrimination and identification of voicing and place contrasts in aphasic patients. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *37*(1), 107.