Individual differences in lexical contributions to speech perception

Nikole Giovannone[a,b] and Rachel M. Theodore [a,b]

[a]Department of Speech, Language, and Hearing Sciences
University of Connecticut
2 Alethia Drive, Unit 1085
Storrs, CT 06269-1085
United States

[b]Connecticut Institute for the Brain and Cognitive Sciences
University of Connecticut
337 Mansfield Road, Unit 1272
Storrs, CT 06269-1272
United States

There are no conflicts of interest to declare. This work was supported by NIH NIDCD grant R21DC016141 RMT, NSF grants DGE-1747486 and DGE-1144399 to the University of Connecticut, and by the Jorgensen Fellowship (University of Connecticut) to NG.

Author to whom correspondence should be addressed:
Rachel M. Theodore, rachel.theodore@uconn.edu

Abstract

Purpose: The extant literature suggests that individual differences in speech perception can be linked to broad receptive language phenotype. For example, a recent study found that individuals with a smaller receptive vocabulary showed diminished lexically guided perceptual learning compared to individuals with a larger receptive vocabulary. Here we examined (1) whether such individual differences stem from variation in reliance on lexical information or variation in perceptual learning itself, and (2) whether a relationship exists between between lexical recruitment and lexically guided perceptual learning more broadly, as predicted by current models of lexically guided perceptual learning.

Method: In Experiment 1, adult participants (*n* = 70) completed measures of receptive and expressive language ability, lexical recruitment, and lexically guided perceptual learning. In Experiment 2, adult participants (*n* = 120) completed the same lexical recruitment and lexically guided perceptual learning tasks to provide a high-powered replication of the primary findings from Experiment 1.

Results: In Experiment 1, individuals with weaker receptive language ability showed *increased* lexical recruitment relative to individuals with higher receptive language ability; however, receptive language ability did not predict the magnitude of lexically guided perceptual learning. Moreover, the results of both experiments converged to show no evidence indicating a relationship between lexical recruitment and lexically guided perceptual learning.

Conclusions: The current findings suggest that (1) individuals with weaker language ability demonstrate increased reliance on lexical information for speech perception compared to those with stronger receptive language ability, (2) individuals with weaker language ability maintain an intact perceptual learning mechanism, and (3), to the degree that the measures used here accurately capture individual differences in lexical recruitment and lexically guided perceptual learning, there is no graded relationship between these two constructs.

Introduction

In speech perception, listeners must accommodate for the fact that there is no one-to-one mapping between speech acoustics and any given consonant or vowel. Despite this lack of invariance, phonemes are perceived categorically (Liberman et al., 1957) and their representations exhibit a rich internal structure that reflects typicality of speech input (Miller, 1994). The mapping between speech acoustics and speech sounds can be dynamically modified by both bottom-up (Clayards et al., 2008; Kleinschmidt & Jaeger, 2015) and top-down learning mechanisms (Ganong, 1980; Norris et al., 2003).

Indeed, it has long been known that listeners use lexical information to facilitate speech perception (Ganong, 1980). When presented with a potentially ambiguous acoustic variant such as a voice-onset-time (VOT) value ambiguous between /g/ and /k/, listeners are more likely to perceive the variant as a member of the category that is consistent with lexical knowledge (Ganong, 1980). For example, when the variant precedes /ɪs/, listeners are more likely to perceive the variant as /k/ than /g/, consistent with the interpretation that yields the real word *kiss* as opposed to the nonword *giss*. However, when the same variant precedes /ɪft/, listeners are more likely to perceive the variant as /g/, as *gift* is a real word and *kift* is a nonword.

This lexical influence on speech perception (also known as the Ganong effect) can be exploited for lexically guided perceptual learning (Norris et al., 2003; Samuel & Kraljic, 2009), in which repeated exposure to ambiguous input in lexically biasing contexts leads to persistent changes in the mapping between acoustics and speech sounds, even when lexical context is subsequently removed. For example, after repeated exposure to an ambiguous fricative (i.e., spectral energy ambiguous between /s/ and /ʃ/) in place of /s/ in lexical contexts (e.g., in place of /s/ in *pencil*), individuals will categorize a continuum of sounds ranging from /s/ to /ʃ/ as having more /s/ than /ʃ/ tokens. However, if individuals instead receive exposure to the ambiguous fricative in place of /ʃ/ in lexical contexts (e.g., in place of /ʃ/ in *ambition*), then they will categorize the same continuum of sounds as having more /ʃ/ than /s/ tokens. Thus, lexically guided perceptual learning allows listeners to dynamically modify the mapping between speech acoustics and speech sound categories, even when disambiguating lexical context is subsequently removed. Learning in this paradigm is robust; it extends beyond the boundary region to facilitate a comprehensive reorganization of phonetic category structure (Drouin et al., 2016; Xie et al., 2017). Moreover, learning can persist over time (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). Lexically guided perceptual learning is often assessed using a between-subjects design in which one group of listeners receives an exposure block biased towards /s/ perception followed by a test block, while the other receives an exposure block biased towards /ʃ/ perception before test. In the absence of any additional input from exposure talker, learning can be observed following both short and long delays between exposure and test (Kraljic & Samuel, 2005; Eisner & McQueen, 2006). However, if listeners hear a second, opposite exposure block from the same speaker, listeners can rapidly retune to the talker's new input, as has been shown when lexically guided perceptual learning is assessed using a within-subjects design (Saltzman & Myers, 2018).[1]

Though lexically guided perceptual learning is a robust phenomenon when assessed at the group level, individual differences in the degree to which adults learn have been observed. A growing body of research suggests that individual differences in lexically guided perceptual learning may reflect individual variation in the relative weighting of phonetic and lexical

---

[1] A retraction notice (Saltzman & Myers, 2020) for this study was issued after the initial submission of the current manuscript. Because the results presented in Saltzman and Myers (2018) contributed to the scientific premise of the current work, we describe them here so that the introduction is a veridical representation of our understanding of the scientific record as this study was developed.

information for speech perception. For example, Scharenborg, Weber, and Janse (2015) found that older adults with higher attention-switching capability showed *decreased* lexically guided perceptual learning compared to those with lower attention-switching capability. Attention-switching was measured via the Trail-Making Test, a standardized measure in which participants must connect alternating letters and numbers in sequence (Reitan, 1958). The authors suggested that this finding may reflect individuals with higher attention-switching ability relying more on phonetic information, whereas individuals with lower attention-switching ability instead rely more on lexical information. Because this study tested older adults, future research is needed in order to determine whether such relationships will also be observed in different populations.

Recent findings from Colby, Clayards, and Baum (2018) lend additional support to the hypothesis that differences in lexical access contribute to individual differences in lexically guided perceptual learning. Specifically, individuals with a lower receptive vocabulary showed diminished lexically guided perceptual learning compared to individuals with a higher receptive vocabulary. Colby and colleagues assessed individual differences in both distributional learning and lexically guided perceptual learning in two age groups (younger and older adults) using a between-subjects design. In both age groups, receptive language ability (as measured by the Peabody Picture Vocabulary Test [PPVT-3]; Dunn & Dunn, 1997) predicted perceptual learning such that individuals with lower PPVT scores demonstrated less learning-consistent responses in both the distributional learning and lexically guided perceptual learning tasks. Similar to the hypothesis of Scharenborg and colleagues (2015), Colby et al. (2018) suggested that this pattern may reflect individuals with a larger vocabulary relying on lexical information to a greater degree than those with a smaller vocabulary.

While the hypothesis that lexical recruitment modulates lexically guided perceptual learning was not directly tested in these studies, it is consistent with work demonstrating that individuals do in fact differ in the degree to which they rely on lexical information for speech perception (Ishida et al., 2016). Ishida and colleagues tested listeners on two tasks. In the first task, listeners made same-different judgements for pairs of stimuli consisting of a natural speech token and a locally time-reversed speech token. Stimuli consisted of both word and nonword items, and the lexical effect was quantified as the difference in sensitivity (d') between word and nonword items. The second task was phonemic restoration, in which listeners judged the phonetic similarity between acoustically modified and unmodified versions of word and nonword items. In the phonemic restoration task, the lexical effect was quantified as the difference in phonemic restoration between word and nonword items. Ishida and colleagues found that (1) individuals varied in the degree to which lexical status influenced performance and (2) individual differences in lexical reliance were stable across tasks such that individuals who showed a stronger effect of lexical status on the perceived intelligibility of locally time-reversed speech stimuli also showed a stronger effect of lexical status on phonemic restoration. These findings suggest that some adults rely more heavily on lexical information than others, and that lexical reliance is not dependent on a specific task.

Moreover, the hypothesis that lexical recruitment is directly related to lexically guided perceptual learning (Colby et al., 2018; Scharenborg et al., 2015) is consistent with current leading models of speech perception that account for lexically guided perceptual learning. Two classes of speech perception models have accounted for the process by which lexical information influences speech perception, as shown in Figure 1. Interactive theories such as TRACE (McClelland & Elman, 1986) posit that lexical information guides perception online – feedback from the lexicon can influence perception itself. In the TRACE model, acoustic input activates feature information, which feeds forward to the phoneme level and then to the lexical level. A defining aspect of the TRACE architecture is that activation can feed backward from the lexical level to the phoneme level and from the phoneme level to the feature level. Phonemic decisions, such as identifying which of two phonemes are heard during the standard lexically guided perceptual learning test task, are modeled as the node with the highest level of activation in the phoneme layer. A Hebbian learning dynamic in the TRACE model allows lexical feedback to strengthen the bidirectional connections between lexical, phoneme, and feature levels based on prior exposure, leading to perceptual learning (i.e., an adjusted connection between the initially ambiguous input and phonemes) even in nonword contexts (Mirman et al., 2006). The TRACE model suggests that lexical recruitment, which is often measured in the form of lexical effects (such as the Ganong effect), necessarily contribute to the phenomenon of lexically guided perceptual learning (Mirman et al., 2006).

Modular (i.e., feed-forward) theories such as Merge (Norris et al., 2000) posit that lexical information does not modify online processing, but instead guides processing at a later decision-level stage. In the Merge model, acoustic input activates nodes at a prelexical (phoneme) level, which feeds forward to the lexical level to facilitate word recognition. Unlike TRACE, there is no feedback from the lexical level to earlier processing levels. To model phonemic decisions, Merge posits that information from both the phoneme and lexical levels feeds to separate decision nodes, which are responsible for determining phonetic categorization. In this way, the decisions made during speech perception are influenced by both phonemic and lexical information, but without lexical information directly feeding back to the phonemic level. Learning in this model occurs when activation from the phoneme and lexical levels is mismatched at the decision level. As a result, a training signal from the decision level modifies prelexical representations, thus modeling a learning effect that can generalize across words. As in TRACE, learning is contingent on lexical recruitment in the Merge model.

Though these models differ in whether lexical information directly feeds back to phonemic representations, they converge on three points for modeling individual differences in lexically guided perceptual learning. First, activation of units within the phonemic and lexical levels is probabilistic, meaning that a specific phoneme/word may be activated with high probability (e.g., 0.9) while other phonemes/words may also be activated for the same input but with a low probability (e.g., 0.1). Probabilistic activation of representational units is fully consistent a wide body of literature for spoken word recognition (e.g., Allopenna et al., 1998; McClelland & Elman, 1986). Second, both models posit that lexically guided perceptual learning cannot occur without lexical activation, an assumption that is supported by findings demonstrating that lexically guided perceptual learning does not occur when exposure consists of ambiguous sounds embedded in nonwords (Norris et al., 2003). Third, both models dissociate online lexical processing from learning within their architectures. That is, though lexical recruitment is necessary for lexically guided perceptual learning to occur, it is not sufficient for learning; lexical information needs to be passed to an intact learning mechanism that modifies the mapping between acoustics and phonemes.

Within these frameworks, individual differences in lexically guided perceptual learning can be modeled in at least two ways. First, individual differences in learning could be accommodated by positing that the degree to which information from the lexical level contributes to reaching a lexical decision (whether online or post-perceptually) has the potential to vary on an individual level. Such differences may reflect the relative availability of acoustic and lexical information, leading some individuals to weight acoustic information more highly than lexical information, or vice versa. These differences may then feed into the learning mechanism, influencing the degree to which individuals dynamically adapt to variation in the speech signal. Second, individual differences in learning could be modeled by variability in the learning mechanism itself. Specifically, both models allow for the possibility that the learning mechanism itself can be selectively impaired, without impairment in lexical recruitment; thus, an individual who demonstrates strong lexical recruitment (for example, a strong Ganong effect) may not necessarily demonstrate an equivalently strong learning effect. Recall that Colby and colleagues (2018) hypothesized that lexically guided perceptual learning was diminshed in those with weaker receptive language due to weaker use of lexical information during speech perception. This hypothesis is fully consistent with the models described above, but is potentially at odds with findings examining lexical recuitment in children with specific language impairment (SLI). Schwartz et al. (2013) measured the magnitude of the Ganong effect in children with and without SLI and observed a larger Ganong effect in children with SLI compared to their typically developing peers. This finding suggests that individuals with weaker receptive language ability may show *increased* reliance on lexical information for speech perception, in opposition to Colby and colleagues' suggestion that weaker receptive language ability is associated with *decreased* reliance on the lexicon. While these seemingly contrary findings may have arisen from any of the methodological differences between these studies, it is theoretically possible that weaker receptive language ability can be associated with both an increased reliance on the lexicon *and* a deficit in lexically guided perceptual learning created by impairment to the learning mechanism. Consistent with the TRACE and Merge frameworks, an impaired learning mechanism in individuals with weaker receptive language ability would result in deficits to lexically guided perceptual learning despite intact (or even stronger) lexical recruitment.

Within this context, the goal of the current investigation is twofold. First, we examine whether the relationship between receptive language ability and lexically guided perceptual learning can be attributed to individual differences in lexical reliance in individuals with lower language ability, or whether they are attributable to variation in the learning mechanism itself. Second, we examine whether there is a relationship between lexical recruitment and lexically guided perceptual learning, as is predicted by both the TRACE and Merge models. To do so, participants in Experiment 1 completed four subtests of the Clinical Evaluation of Language Fundamentals – 5th Edition (CELF; Wiig, Semel, & Secord, 2013) in addition to tasks assessing the Ganong effect and lexically guided perceptual learning. CELF subtests consisted of two that assess expressive language (Formulated Sentences, Recalling Sentences) and two that assess receptive language (Understanding Spoken Paragraphs, Semantic Relationships). Separate expressive and receptive language profiles were obtained in order to examine potential specificity of these constructs as contributors to individual differences in lexical recruitment and perceptual learning; two measures for expressive and receptive language were collected in order to assess convergence in results between the measures assessing each of these broad constructs. Participants in Experiment 2 completed the same Ganong and lexically guided perceptual leaning tasks as for Experiment 1 but did not complete the CELF measures.

If individual variation in lexically guided perceptual learning is due to individual differences in lexical recruitment, then individuals with weaker receptive language ability should show a diminished Ganong effect in addition to diminished perceptual learning. Alternatively, if weaker perceptual learning in individuals with weaker receptive langauge abiltiy is attributable to impairment in the learning mechanism itself, consistent with the procedural deficit (Ullman & Pierpont, 2005) and statistical learning deficit (Hsu & Bishop, 2014) hypotheses of language impairment, then individuals with weaker language ability will show diminished lexically guided perceptual learning regardless of the degree of lexical recruitment. Independent of language ability, if lexical recruitment modulates lexically guided perceptual learning – as predicted by both the TRACE and Merge models, then performance on the Ganong task will predict performance on the lexically guided perceptual learning task such that increased lexical recruitment is associated with increased perceptual learning.

Experiment 1

Method

Participants. The participants were 70 native speakers of American English (20 men, 50 women) between 18 and 26 years of age (mean = 20, SD = 2) who were recruited from the University of Connecticut community. Thirty-one participants had experience with a second language, with self-reported proficiency of novice (*n* = 18), intermediate (*n* = 11), or advanced (*n* = 2). All participants passed a pure tone hearing screen administered at 25 dB for octave frequencies between 500 and 4000 Hz and had nonverbal intelligence within normal limits (range = 86 – 122, mean = 103, SD = 9) as assessed using the standard score of the Test of Nonverbal Intelligence – 4th Edition (TONI; Brown, Sherbenou, & Johnsen, 2010). The TONI is normed to reflect a population mean of 100 (SD = 15). All participants completed Ganong and lexically guided perceptual learning tasks (described below) in addition to assessments of expressive and receptive language ability. Language ability was assessed using the standard score[2] of four subtests from the Clinical Evaluation of Language Fundamentals – 5th Edition

---

[2] Twelve of the 70 participants were beyond the oldest age (21 years) provided for the standard score conversion of the CELF-5. Calculation of standard scores for these participants was made using the oldest age provided for the conversion, which is sensible given that this age bracket represents a maturational end-state. However, all analyses conducted with standard scores were also conducted with raw scores, and qualitatively similar results were observed in all cases.

(CELF; Wiig et al., 2013); scoring (e.g., trial-level scoring, calculation of standard score) was performed as outlined in the administration manual.

Expressive language was assessed using the Formulated Sentences and Recalling Sentences subtests. For Formulated Sentences, participants are asked to generate a sentence to describe a specific picture that contains one (or two) words provided by the experimenter. Responses are scored based on the appropriateness of the sentence in the context of the stimulus picture. For Recalling Sentences, participants are required to repeat verbatim a sentence provided by the experimenter. Though the Recalling Sentences task requires contributions from perception and memory in order to be completed successfully, this subtest is characterized as an expressive language measure in the CELF manual. Receptive language was assessed using the Understanding Spoken Paragraphs and Semantic Relationships subtests. For Understanding Spoken Paragraphs, participants hear a series of short passages read by the experimenter and answer comprehension questions for each passage. For Semantic Relationships, participants are asked to solve short word problems that probe semantic knowledge by selecting the two correct items from a set of four items following a spoken prompt. An example problem is hearing “Jan saw Pedro. Pedro saw Francis. Who was seen?” and being shown *Jan*, *Dwayne*, *Pedro*, and *Francis* as possible response items (with the correct answers being *Pedro* and *Francis*). Due to an error in implementing the reversal rule during CELF-5 administration, the number of participants that could be accurately scored for a given subtest varied slightly across the four subtests (Formulated Sentences, *n* = 54; Recalling Sentences; *n* = 58; Understanding Spoken

---

These analyses can be viewed by executing the supplemental analysis script provided in the OSF repository: https://osf.io/r5sp9/.

Paragraphs, *n* = 70; Semantic Relationships, *n* = 63).[3] Figure 2 shows the distribution of standard scores for each of the four CELF subtests; standard scores for the CELF subtests reflect a population mean of 10 (SD = 3).

Stimuli: Ganong task. Stimuli for the Ganong task were two eight-step voice-onset-time (VOT) continua that perceptually ranged from *giss* to *kiss* and *gift* to *kift*, respectively. Both continua were created using the Praat software (Boersma, 2002) from tokens produced by a native male speaker of American English. Drawing from recorded productions that were free of acoustic artifact, a single /ɪs/ portion was selected for the *giss*-*kiss* continuum and a single /ɪft/ portion was selected for the *gift*-*kift* continuum such that duration of the /ɪs/ (374 ms) and /ɪft/ (371 ms) portions were equivalent. To create the VOT portion (cueing the initial consonant), eight different VOTs (17, 21, 27, 37, 46, 51, 59, and 71 ms) were created by successively removing energy from the aspiration region of a natural *kiss* production. The first step contained the burst plus the first quasi-periodic pitch period; subsequent steps contained this burst in addition to aspiration energy that increased across continuum steps. These eight VOTs were then spliced to the selected /ɪs/ and /ɪft/ portions. With this procedure, the only difference among steps within a given continuum was VOT duration and the only difference between continua for a given step was lexical context (cued by the /ɪs/ or /ɪft/ context). All stimuli were normalized for

---

[3] The nature of the administration error for the Formulated Sentences, Recalling Sentences, and Semantic Relationships subtests, discovered after data collection was completed in conjunction with a double-check of the scoring data, is as follows. For all three subtests, participants began the test at the item appropriate for their age. In some cases, the participant did not meet the reversal rule (i.e., perfect score on the first two consecutive items) and the administrator failed to go back to the first item and test forward to the initial start point. As a consequence, the raw score for the affected participants may be higher than what would have been obtained if the reversal rule had been implemented correctly. As described in the main text, affected participants were removed from specific subtest analyses in light of this error. We note that even with their removal, the sample size for each subtest analysis ($n \geq 54$ in all cases) remains large relative to similar recent investigations (e.g., $n = 31$ for the young adult sample in Colby et al., 2018).

peak amplitude.

As described above, the VOT portion cueing the initial consonant was identical between the two continua, with the coda portion (i.e., /s/ and /ft/) providing the critical lexical context required to elicit a Ganong effect. A preliminary experiment was conducted in order to ensure that perception of VOT was indeed equivalent between the two continua in the absence of lexical context. This study was hosted online by the Gorilla platform following procedures described for Experiment 2. Participants (n = 20 monolingual English speakers between 20 and 34 years of age with no history of language disorders) were recruited from the Prolific participant pool and passed the headphone screen of Woods and colleagues (Woods, Siegel, & McDermott, 2017) at the beginning of the experiment. Stimuli for this preliminary study consisted of those described for the Ganong task (2 continua x 8 steps) in addition to two parallel continua that were created by removing the coda portion from each of the 16 stimuli, thus creating two "control" continua that each perceptually ranged from /gɪ/ to /kɪ/. All participants completed two blocks of phonetic categorization, one for the control stimuli and one for the Ganong stimuli; block order was fixed across participants (control block followed by Ganong block). Each block consisted of 10 repetitions of the 16 stimuli appropriate for each block presented in randomized order. On each trial, participants were asked to identify the initial sound as either /g/ or /k/ by pressing an appropriately labeled key on the keyboard.

Mean proportion /k/ responses for each continuum in each block are shown in Figure 3. Visual inspection suggests that a Ganong effect was indeed observed in the Ganong block, reflecting more /k/ responses for the *giss-kiss* continuum compared to the *gift-kift* continuum. In contrast, /k/ responses appear equivalent between the two continua in the control block.

To confirm this pattern statistically, trial-level responses (/g/ = 0, /k/ = 1) were submitted to a generalized linear mixed effects model as implemented in the lme4 (Bates et al., 2015) package in R. The fixed effects included VOT (scaled/centered around the mean), continuum, block, and all interactions. Continuum and block were sum-coded (continuum: *gift-kift* = -0.5, *giss-kiss* = 0.5; block: control = -0.5, Ganong = 0.5). The random effects structure consisted of random intercepts by participant and random slopes for VOT, continuum, and block by participant. An interaction between continuum and block was observed ($\hat{\beta} = 2.248$, $SE = 0.174$, $z = 12.927$, $p < .001$). To explicate the nature of the interaction, separate models with the fixed effects of VOT, continuum, and their interaction were constructed for each block. The main effect of continuum was significant in the Ganong block ($\hat{\beta} = 1.836$, $SE = 0.467$, $z = 3.936$, $p < .001$) but not in the control block ($\hat{\beta} = -0.175$, $SE = 0.159$, $z = -1.100$, $p = .271$). This preliminary study confirms that the stimuli developed for the Ganong task are appropriate for use in the primary experiment.

Stimuli: Perceptual learning task. Stimuli for the lexically guided perceptual learning task were those in Myers and Mesite (2014) to which the reader is referred for comprehensive details on stimulus creation. We used this stimulus set given that it has been shown to successfully elicit lexically guided perceptual learning across numerous samples (Drouin et al., 2016; Drouin & Theodore, 2018; Myers & Mesite, 2014; Saltzman & Myers, 2018). In brief, there were two sets of exposure stimuli (one for the /s/-bias block and one for the /ʃ/-bias block) and one set of test stimuli, all produced by a single female native speaker of American English. The exposure sets each consisted of 200 auditory items (100 words and 100 nonwords). For word items, 20 were critical /s/ items (e.g., *pencil*), 20 were critical /ʃ/ items (e.g., *ambition*), and 60 were filler items that contained no instances of /s/ or /ʃ/. For the /s/-bias set, the medial /s/ of the critical /s/ items was replaced with an ambiguous fricative (consisting of a 50:50 blend of /s/ and

/ʃ/ sounds). For the /ʃ/-bias set, the medial /ʃ/ of the critical /ʃ/ items was replaced with an ambiguous fricative (i.e., a 50:50 blend of /s/ and /ʃ/ sounds). Test stimuli consisted of a seven-step continuum that perceptually ranged from *shine* to *sign*. The continuum was created by blending the initial fricatives from natural productions of *sign* and *shine* in different proportions ranging from 20:80 (20% /s/ and 80% /ʃ/, the *shine* end of the continuum) to 80:20 (80% /s/ and 20% /ʃ/, the *sign* end of the continuum) in 10% steps. All stimuli were normalized for peak amplitude using Praat (Boersma, 2002).

Procedure. Participants were tested individually in a sound-attenuated booth. Stimuli were presented via headphones (Sony MDR-7506) at a comfortable listening level held constant across participants. Responses were made via button box (Cedrus RB-740). Stimulus presentation and response collection were controlled using SuperLab (version 4.5) running on a Mac OS X operating system. For both tasks, participants were directed to respond as quickly and accurately as possible and to guess if they were unsure.

The Ganong task consisted of 160 trials, formed by 10 repetitions of the eight continuum steps for each of the *giss–kiss* and *gift–kift* continua; items were presented in randomized order (ISI = 1500 ms). On each trial, participants indicated whether the initial sound was either /g/ or /k/. Participants then completed two blocks of lexically guided perceptual learning. All participants first received /s/-bias exposure (followed by test) and then received /ʃ/-bias exposure (followed by test). During exposure, the 200 items appropriate for the specific exposure block were presented in randomized order (ISI = 2000 ms). On each trial, participants indicated whether each item was a word or nonword. During test, eight repetitions of the seven test stimuli were presented in randomized order (ISI = 2000 ms); participants were asked to categorize each item as either *sign* or *shine*.

All participants completed the Ganong task before the lexically guided perceptual learning task in order to mitigate the possibility that the Ganong effect would be inflated due to possible carryover effects from the LGPL task. Specifically, the LGPL task requires listeners to make lexical decisions during the exposure phase, but the Ganong task requires listeners to make phonetic decisions. If listeners had completed the LGPL task first, then they may have been primed to approach the Ganong task as a lexical decision task instead of a phonetic categorization task. Participants were given a brief break in between the two tasks and received monetary compensation or partial course credit for their participation.

Results

Ganong task. Trial-level data and a script (in R) to reproduce all analyses presented in this manuscript can be retrieved at: https://osf.io/r5sp9/. Responses on the Ganong task were coded as either /g/ (0) or /k/ (1). Trials for which no response was provided were excluded (< 1% of the total trials). To visualize performance in the aggregate, mean proportion /k/ responses was calculated for each participant for each step of the two continua. Responses were then averaged across participants and are shown Figure 4, panel A. Visual inspection of this figure reveals a robust Ganong effect; more /k/ responses are observed for the *giss-kiss* continuum compared to the *gift-kift* continuum.

To examine this pattern statistically, trial-level responses (0 = /g/, 1 = /k/) were fit to a generalized linear mixed-effects model (GLMM) using the glmer() function with the binomial response family (i.e., a logistic regression) as implemented in the lme4 package (Bates et al., 2015) in R. The fixed effects included VOT, continuum, and their interaction. VOT was entered into the model as continuous variable, scaled and centered around the mean. Continuum was sum-coded (*giss–kiss* = 0.5, *gift–kift* = -0.5). The random effects structure consisted of random intercepts by participant and random slopes by participant for VOT, continuum, and their interaction. As expected, the model showed a significant effect of VOT ($\hat{\beta}$ = 3.463, *SE* = 0.153, *z* = 22.698, *p* < .001), indicating that /k/ responses increased as VOT increased. There was a significant effect of continuum ($\hat{\beta}$ = 1.265, *SE* = 0.177, *z* = 7.141, *p* < .001), with the direction of the beta estimate indicating increased /k/ responses in the *giss–kiss* compared to the *gift–kift* continuum. There was also an interaction between continuum and VOT ($\hat{\beta}$ = 1.097, *SE* = 0.187, *z* = 5.882, *p* < .001), indicating that the magnitude of the Ganong effect was not equivalent across continuum steps. Thus, the results of this model confirm the presence of a Ganong effect for participants in the aggregate.

The next set of analyses were conducted in order to examine whether the magnitude of the Ganong effect was linked to performance on the receptive and expressive language measures. To do so, trial-level data (0 = /g/ response, 1 = /k/ response) were fit to a series of mixed effects models, one for each CELF subtest. Subtests were tested in separate models due to collinearity among predictors. The fixed effects in each model consisted of VOT, continuum, the CELF subtest, and all interactions among the three factors. Continuum was sum-coded as described for the aggregate model; VOT and CELF subtest were entered into the model as continuous variables (scaled/centered around the mean). The random effects structure consisted of random intercepts by participant and random slopes by participant for VOT, continuum, and their interaction. In all models, evidence of a link between subtest performance and the Ganong effect would manifest as an interaction between continuum and subtest.

The results of the four models are shown in Table 1. There was no interaction between continuum and subtest for Formulated Sentences (*p* = .911), an expressive language measure. However, the continuum by subtest interaction was significant for the expressive Recalling

Sentences measure (*p* = .020) and both receptive language measures (Understanding Spoken Paragraphs, *p* = .008; Semantic Relationships, *p* = .007). The (negative) direction of the beta estimate for the significant interactions indicates a larger Ganong effect (i.e., difference between the *giss*-*kiss* and *gift*-*kift* continua) for those with *weaker* receptive language scores.

Figure 5, panel A shows the beta estimate and 95% confidence interval for the continuum by subtest interaction for all four subtests. To illustrate the nature of the interaction, Figure 5, panel B shows performance on the Ganong task according to a median split of participants based on USP score; though both groups show a Ganong effect, the magnitude of this effect is larger in those with weaker receptive language as indexed by USP score. The same qualitative pattern – a larger Ganong effect for those with weaker compared to stronger language scores – was present for the other two significant interactions (i.e., Recalling Sentences, Semantic Relationships) as indicated by the negative beta estimate for each of the interaction terms.

In addition, single-order correlations between the magnitude of the Ganong effect (quantified as the difference in proportion /k/ responses between the *giss*-*kiss* and *gift*-*kift* continua) and subtest standard scores were run to facilitate comparison with the extant literature. These results are presented in Table 2. In all cases, qualitatively similar results to those demonstrated by the GLMM analyses were found.

Perceptual learning task. Accuracy (proportion correct) on the lexical decision task during the exposure phase was near ceiling (mean = 0.95, SD = 0.03, range = 0.86 – 0.99).[4]

---

[4] Recall that Colby et al. (2018) found that individuals with stronger receptive vocabulary (as measured by the PPVT) showed increased lexically guided perceptual learning compared to those with weaker receptive vocabulary. Though Experiment 1 did not include a standardized measure of receptive vocabulary, performance on the lexical decision task provides an indirect measure of vocabulary. A series of exploratory analyses was conducted for the lexical decision data. Mean accuracy during exposure was not correlated with any of the four CELF subtests, and mean accuracy during exposure was not a significant predictor of the magnitude of perceptual

Responses at test for the perceptual learning task were coded as either /ʃ/ (0) or /s/ (1). Trials for which no response was provided were excluded (< 1% of the total trials). To visualize performance in the aggregate (Figure 6, panel A), mean proportion *sign* responses was first calculated by participant in each half of the two test blocks (block 1 = /s/-bias, block 2 = /ʃ/-bias) at each step of the test continuum. Performance at test is considered over time (i.e., first half vs. second half) given research showing that lexically guided perceptual learning is attenuated throughout the test period (Liu & Jaeger, 2018, 2019). That is, recent findings have shown that exposure to the flat frequency distributions at test (e.g., eight repetitions of each of the seven test stimuli) promotes unlearning of the biased input during exposure presumably due to distributional learning that occurs throughout the test period (Liu & Jaeger, 2018, 2019). Indeed, visual inspection of Figure 6, panel A suggests that the lexically guided perceptual learning effect is present in the first half of the test block, but attenuated in the second half of the test block.

To examine these patterns statistically, trial-level responses (0 = /ʃ/, 1 = /s/) were fit to a GLMM. The fixed effects included step, bias, half, and all interactions between the three factors. Step was entered into the model as a continuous variable (scaled/centered around the mean). Bias and half were sum-coded (/s/-bias = 0.5, /ʃ/-bias = -0.5; first half = 0.5, second half = -0.5). The

---

learning during the test phase. When accuracy on the lexical decision task was measured separately for the four item types presented during exposure (i.e., critical /s/ words, critical /ʃ/ words, filler words, nonwords), there was (1) no correlation between any of the CELF subtests and accuracy for /s/ words or /ʃ/ words, (2) significant, moderate positive correlations between the two receptive language subtests and accuracy for filler words, and (3) a significant but weak positive correlation between the Semantic Relationships subtest and accuracy for nonwords. However, accuracy for either filler words or nonwords was not a significant predictor of the magnitude of perceptual learning during the test phase. These exploratory analyses can be viewed in the supplementary analysis script on the OSF repository for this manuscript: https://osf.io/r5sp9/.

random effects structure consisted of random intercepts by participant and random slopes by participant for step, bias, and half. The model showed a main effect of step ($\hat{\beta}$ = 3.899, *SE* = 0.179, *z* = 21.835, *p* < .001), with /s/ responses increasing across the test continuum. There was also a main effect of bias ($\hat{\beta}$ = 0.503, *SE* = 0.191, *z* = 2.630, *p* = .009), with more /s/ responses in the /s/-bias block compared to the /ʃ/-bias block, indicative of lexically guided perceptual learning. However, there was a significant interaction between bias and half ($\hat{\beta}$ = 0.913, *SE* = 0.232 *z* = 3.938, *p* < .001). Simple slopes analyses showed a robust effect of bias in the first half of the test block ($\hat{\beta}$ = 0.959, *SE* = 0.221, *z* = 4.338, *p* < .001), but no effect of bias in the second half of the test block ($\hat{\beta}$ = 0.046, *SE* = 0.226, *z* = 0.204, *p* = .839). Thus, a robust perceptual learning effect is observed at test, but it is limited to the first half of the test period in the current data, consistent with research showing that learning in this paradigm is attenuated throughout the test block as a consequence of exposure to the flat frequency distributions presented at test (Liu & Jaeger, 2018, 2019).[5]

Given that perceptual learning in the aggregate was only observed during the first half of the test period, consistent with past research (Liu & Jaeger, 2018, 2019) – and that past research has shown that receptive language ability is linked to distributional learning (Colby et al., 2018; Theodore et al., 2019), the presumed mechanism responsible for diminished learning during the lexically guided perceptual learning test phase – the next set of analyses tested for links between the language measures and perceptual learning isolating performance to the first half of each test

---

[5] For both Experiment 1 and Experiment 2, an additional analysis examined performance across consecutive test block halves (e.g., first half of /s/-bias test block, second half of /s/-bias test block, first half of /ʃ/-bias test block, second half of /ʃ/-bias test block) using sliding contrast comparisons. The results of these analyses suggest that performance in the second block reflected boundary movement in the opposite direction of the first block instead of simple “unlearning” during the first test block. These analyses can be viewed on the OSF repository for this manuscript: https://osf.io/r5sp9/.

block. To do so, trial-level data (0 = /ʃ/ response, 1 = /s/ response) were fit to a series of mixed effects models, one for each CELF subtest. Subtests were tested in separate models due to potential collinearity among predictors. The fixed effects in each model consisted of step, bias, the CELF subtest, and all interactions among the three factors. Bias was sum-coded as described for the aggregate model; step and CELF subtest were entered into the model as continuous variables (scaled/centered around the mean). The random effects structure consisted of random intercepts by participant and random slopes by participant for step, continuum, and their interaction.

In all models, evidence of a link between subtest performance and the perceptual learning effect would manifest as an interaction between bias and subtest. The results of the four mixed effects models are shown in Table 3. There was no significant interaction between bias and subtest for any of the expressive or receptive language measures.

Figure 7, panel A shows the beta estimate and 95% confidence interval for the bias by subtest interaction for all four subtests. To illustrate the nature of the (null) interactions, Figure 7, panel B shows performance on the perceptual learning task according to a median split of participants based on USP score. Though there is a numerical trend for the learning effect to be larger for those with weaker compared to stronger receptive language (i.e., larger beta estimates for the bias by subtest interactions for the two receptive language measures compared to the expressive language measures), these relationships were not statistically reliable.

Like for the Ganong task, single-order correlations between the magnitude of the learning effect (quantified as the difference in proportion /s/ responses between the first half of the /s/-bias and /ʃ/-bias test blocks) and subtest standard scores were run to facilitate comparison with extant literature. These results are presented in Table 2. In all cases, qualitatively similar results to those demonstrated by the GLMM analyses were found.

Relationship between Ganong and perceptual learning tasks. The results presented thus far show that individuals with weaker receptive language showed a larger Ganong effect, consistent with past research (Schwartz et al., 2013). The same pattern also held for one of the expressive language measures. However, none of the language measures was a reliable predictor of the magnitude of the perceptual learning effect. This finding contrasts with results from Colby and colleagues (Colby et al., 2018), who found that stronger receptive vocabulary was associated with increased learning. A final analysis tested the prediction from both modular and interactive accounts of perceptual learning (Figure 1), which posit a positive relationship between lexical recruitment and strength of perceptual learning. For each participant, we (1) quantified the magnitude of the Ganong effect as the difference in proportion /k/ responses between the *giss-kiss* and *gift-kift* continua and (2) quantified the magnitude of the perceptual learning effect as the difference in proportion /s/ responses between the first half of the /s/-bias and /ʃ/-bias test blocks. In both cases, higher difference scores indicate larger effects. As can be seen in Figure 8, panel A, there was no correlation between the two measures (r = -0.08, p = .492). This held even when simply comparing the correlation of rank order between the two measures (ρ = -0.07, p = .587).

Experiment 2

The results of Experiment 1 showed that receptive language ability, as measured by the Understanding Spoken Paragraphs and Semantic Relationships CELF subtests, was inversely associated with lexical recruitment. Compared to individuals with stronger receptive language, individuals with weaker receptive language showed *increased* reliance on lexical information. The same relationship was also observed for the Recalling Sentences subtest, which is specified as a measure of expressive language in the CELF manual, but may in fact reflect receptive language ability given that perception and memory processes are required to successfully complete this task. However, none of the CELF subtests was associated with lexically guided perceptual learning, and no reliable relationship in the magnitude of the lexical recruitment and perceptual learning effects was observed. That is, the magnitude of the Ganong effect did not predict the magnitude of the learning effect, in contrast to predictions made by current theories of lexically guided perceptual learning. Moreover, the lack of a relationship between the two tasks challenges previous interpretations of individual differences in lexically guided perceptual learning.

Recall that in Experiment 1, bias in the perceptual learning task was manipulated within-subjects in order to address potential issues with asymmetry in learning across bias conditions. Though past research has indeed found evidence that listeners rapidly recalibrate for lexically guided perceptual learning when bias is manipulated within-subjects (Saltzman & Myers, 2018), this method of measuring perceptual learning remains nonstandard in this domain. In this context, Experiment 2 was conducted as a replication of Saltzman and Myers (2018), who found that listeners rapidly retuned phonetic boundaries when lexically-biased exposure changed within an experimental session. Doing so would confirm the validity of the methodological decision to manipulate bias within-subjects in the current work. In addition, we aimed to replicate the (null) relationship between the magnitude of the Ganong effect and the magnitude of the learning effect that was observed in Experiment 1 with a larger sample size.

Method

Participants. Participants (n = 120; 61 men, 59 women) were recruited from the Prolific participant pool (https://www.prolific.co). All participants were monolingual speakers of

American English between 19 and 35 years of age (mean = 26, SD = 5) who were currently residing in the US and had no history of language disorders according to self-report. All participants passed the headphone screen of Woods et al. (2017), which is a protocol designed to ensure headphone compliance for web-based studies. Participants were compensated with $5.33 for completing the study. An additional 43 participants were tested but excluded from the study due to failure to pass the headphone screen or failure to meet compliance checks (e.g., participant only pressed one button for the entire experiment); this attrition rate is consistent with other web-based studies (Brown et al., 2018; Thomas & Clifford, 2017; Woods et al., 2017)

Stimuli and procedure. The stimuli were identical to those used in Experiment 1. The procedure was identical to that outlined for Experiment 1 with three exceptions. First, all testing was completed online. The experiment was programmed using the Gorilla platform (https://gorilla.sc), which was also used to host the study online. Second, participants were randomly assigned to either the SS-SH order group (n = 60) or the SH-SS order group (n = 60) for the lexically guided perceptual learning task. The SS-SH order group thus provides a replication of the learning task used in Experiment 1, where all listeners received /s/-bias exposure (followed by test) and then /ʃ/-bias exposure (followed by test). The SH-SS order group received the same exposure but first completed the /ʃ/-bias block (exposure followed by test) and then completed the /s/-bias block (exposure followed by test). Third, participants did not complete the CELF battery because it cannot be administered in a web-based format.

Results

Ganong task. Performance was analyzed as outlined for the aggregate model in Experiment 1 and is displayed in Figure 4, panel B. The GLMM showed a significant effect of VOT ($\hat{\beta}$ = 2.874, SE = 0.116, z = 24.752, p < .001), a significant effect of continuum ($\hat{\beta}$ = 1.558, *SE* = 0.128, z = 12.135, p < .001), and a marginal interaction between VOT and continuum ($\hat{\beta}$ = 0.201, *SE* = 0.104, z = 1.922, p = .055). These results confirm the presence of a Ganong effect in Experiment 2.

Perceptual learning task. Accuracy (proportion correct) on the lexical decision exposure task was near ceiling (mean = 0.95, SD = 0.04, range = 0.81 – 0.99). For the initial analysis of the test data, performance was analyzed as outlined for the aggregate model in Experiment 1 (thus collapsing across order groups) and is displayed in Figure 6, panel B. As in Experiment 1, a significant interaction was observed between bias and half ($\hat{\beta}$ = 0.875, *SE* = 0.168, z = 5.225, p < .001), with learning attenuated in the second half of the test period compared to the first half of the test period. Given this interaction – and to optimally promote comparison to Experiment 1 – subsequent analyses were limited to performance in the first half of each test block.

The second analysis directly compared learning between the two order groups. Trial-level responses (/s/ = 0, /ʃ/ = 1) were submitted to a GLMM with the fixed effects of step, bias, order, and their interactions. Step was entered as a continuous variable, bias and order were sum-coded (/s/ = 0.5, /ʃ/ = -0.5; SS-SH = 0.5, SH-SS = -0.5). The random effects structure consisted of random intercepts by participant, and random slopes by participant for step, bias, and their interaction.

The model showed a main effect of step ($\hat{\beta}$ = 4.340, *SE* = 0.187, z = 23.178, p < .001) and bias ($\hat{\beta}$ = 1.787, *SE* = 0.310, z = 5.764, p < .001), and an interaction between step and bias ($\hat{\beta}$ = 1.012, *SE* = 0.338, z = 2.999, p = .003), the latter indicating that the magnitude of the learning effect differed across continuum steps. There was no main effect of order ($\hat{\beta}$ = 0.026, *SE* = 0.377, z = 0.069, p = .945), but there was a significant interaction between bias and order ($\hat{\beta}$ = -0.977, *SE* = 0.492, z = -1.985, p = .047). The three-way interaction between step, bias, and order was not reliable ($\hat{\beta} = 0.154$, $SE = 0.470$, $z = 0.327$, $p = .743$).

Simple slopes analyses were used to explicate the bias by order interaction. For the between-subjects comparisons, there was no reliable difference in /s/ responses between the two order groups for either the /s/-bias test block ($\hat{\beta} = -0.462$, $SE = 0.525$, $z = -0.881$, $p = .378$) or the /ʃ/-bias test block ($\hat{\beta} = 0.514$, $SE = 0.360$, $z = 1.427$, $p = .154$). For the within-subjects comparisons, an effect of bias was present in both the SS-SH order group ($\hat{\beta} = 1.298$, $SE = 0.384$, $z = 3.378$, $p = .001$) and the SH-SS order group ($\hat{\beta} = 2.275$, $SE = 0.406$, $z = 5.597$, $p < .001$); however, the effect size (as measured by the beta estimate) is larger in the latter. Collectively, these results indicate that perceptual learning was present for both order groups, but in contrast to Saltzman and Myers (2018), the magnitude of the learning effect was larger for the SH-SS order group compared to the SS-SH order group.

A third analysis tested the between-subjects learning effect in the first and second test blocks in order to assess whether potential carry-over effects from the first test block influence between-subjects performance in the second test block. Trial-level responses (/s/ = 0, /ʃ/ = 1) were submitted to two separate GLMMs, one for each test block. Both models followed the same structure, which included fixed effects of step, bias, and their interaction. Step was entered as a continuous variable; bias was sum-coded as in the aggregate model. The random effects structure consisted of random intercepts by participant, and random slopes by participant for step. Both models showed a significant main effect of bias (test block 1: $\hat{\beta} = 1.467$, $SE = 0.412$, $z = 3.561$, $p < .001$; test block 2: $\hat{\beta} = 1.338$, $SE = 0.458$, $z = 2.920$, $p = .004$).

Using the beta estimates as a measure of effect size, the magnitude of the bias effect is similar between the two test blocks. To confirm this observation statistically, an additional model was tested combining data from both test blocks that included fixed effects of step, bias, block, and all interactions between the three factors. Step was entered as a continuous variable and bias was sum-coded as in the aggregate model. Block was also sum-coded (test block 1 = 0.5; test block 2 = -0.5). Random effects included random intercepts by participant, and random slopes by participant for step, bias, and block. No interaction between bias and block was observed ($\hat{\beta} = -0.091$, $SE = 0.739$, $z = 0.124$, $p = .901$), thus providing no evidence that the magnitude of the between-subjects bias effect differed between the two test blocks.

Relationship between Ganong and perceptual learning tasks. The magnitude of the Ganong effect and the magnitude of the perceptual learning effect was quantified for each participant as described for Experiment 1. The relationship between the two effects is shown in Figure 8, panel B. As for Experiment 1, there was no correlation between the two tasks in terms of either absolute magnitude ($r = -0.01$, $p = .918$) or rank order ($\rho = -0.10$, $p = .277$). The same patterns held when the correlations were performed within each order group separately (SS-SH: $r = 0.09$, $p = .490$, $\rho = -0.07$, $p = .596$; SH-SS: $r = -0.06$, $p = .630$, $\rho = -0.08$, $p = .553$).

Discussion

Summary

The goal of the current study was twofold. First, we assessed whether individual differences in lexically guided perceptual learning associated with receptive language ability reflect variation in lexical reliance or variation in perceptual learning itself. Second, we assessed whether there is a relationship between lexical recruitment and lexically guided perceptual learning in general, as predicted by both interactive and modular models of perceptual learning.

With regard to the first question, the results of Experiment 1 suggest two key findings. First, weaker language ability was associated with a larger Ganong effect, indicative of increased reliance on lexical information in these individuals. The magnitude of the Ganong effect was predicted by both of measures of receptive language ability and one measure of expressive language ability. These results are consistent with findings demonstrating that children with specific language impairment, which is associated with receptive language deficits, exhibit a larger Ganong effect compared to typically-developing children (Schwartz et al., 2013). Second, we found no evidence of a relationship between our measures of lexically guided perceptual learning and language ability, suggesting that individuals with weaker language ability have an intact perceptual learning mechanism despite their weaknesses in broad language phenotype. These results diverge from those of Colby and colleagues (2018), who found that weaker receptive language ability (as measured by receptive vocabulary) was associated with diminished lexically guided perceptual learning. Results were comparable when derived from single-order correlations, which may yield a more transparent measure of effect size, as well as when derived from generalized linear mixed effects models, which specifically model and thus account for individual differences in the identification response function.

Both experiments offer insight on our second question, which concerned the relationship between lexical recruitment and lexically guided perceptual learning in general. Despite the hypothesized relationship between these two constructs, we observed no evidence to suggest a relationship between lexical recruitment and lexically guided perceptual learning across two experiments that collectively tested 190 participants.

Implications for theory

Previous research (Colby et al., 2018; Scharenborg et al., 2015), as well as both the TRACE and Merge models of speech perception, suggest the existence of a relationship between lexical recruitment and lexically guided perceptual learning. For example, Colby et al. (2018) found that individuals with lower receptive vocabulary showed attenuated lexically guided perceptual learning, which they hypothesized may reflect a decreased reliance on top-down information for speech perception. In addition, Scharenborg et al. (2015) suggested that individuals with lower attention-switching capability demonstrate diminished lexically guided perceptual learning effects because they rely on top-down lexical information to a greater degree than those with higher attention-switching capability, who instead rely more highly on bottom-up phonetic information. In the current work, we found no evidence to suggest that individual variation in lexically guided perceptual learning was linked to receptive or expressive language ability; moreover, we found no evidence to suggest a relationship between lexical recruitment and lexically guided perceptual learning more generally.

Though the interpretation of null results is inherently challenging, the lack of a relationship between lexical recruitment and lexically guided perceptual learning may be treated with some degree of credibility. Zheng and Samuel (2020) outlined three criteria that could mitigate concern in interpreting null results: adequate power, sufficient between-subjects variability, and stable within-subjects performance. The current experiments clearly meet the first two criteria. Experiment 1 tested 70 participants and Experiment 2 tested 120 participants. These sample sizes are well above those generally tested for studies of lexically guided perceptual learning, and post-hoc sensitivity analyses suggest that they were sufficiently powered ($1 - \beta = 0.80$, $\alpha = 0.05$) to detect small to moderate effects ($r = 0.33$ given n = 70, $r = 0.25$ given n = 120). The samples tested in this study yielded substantial between-subjects variability for all tasks as shown in Figure 2 (CELF subtest scores) and Figure 8 (performance in the Ganong and lexically guided perceptual learning tasks). Regarding the third criterion, as noted by Zheng and Samuel (2020), the nature of the lexically guided perceptual learning effect makes it very difficult to properly assess its within-subject stability, which we discuss further below.

As in Colby et al. (2018) and Scharenborg et al. (2015), the current study used the lexically guided perceptual learning paradigm. The discrepancy between past research and the current findings may be related to the specific tasks used to measure individual differences in language ability, lexical recruitment, and/or the specific population being tested. Colby et al. (2018) measured receptive language ability using the PPVT, whereas receptive language ability in the current work was measured using two subtests of the CELF. Accordingly, Colby and colleagues measured receptive vocabulary in isolation, whereas the language measures used in the current study encompass multiple elements of receptive language. It may be the case that individual differences in perceptual learning reflect contributions from vocabulary size that are dissociable from measures that assess receptive language ability more broadly. Further research directed towards dissociating which aspects of language processing are related to lexically guided perceptual learning should be conducted through the use of more specific measures of language ability. In past studies, reliance on lexical information was hypothesized to be the mediator of observed relationships between individual differences on the PPVT (for younger and older adults) or Trail Making task (for older adults) and lexically guided perceptual learning. In the current study, lexical recruitment was directly measured for younger adults using the Ganong task. The Ganong task is widely accepted as reflecting the contribution of lexical information to speech perception (Ishida et al., 2016; Pitt, 1995) and is therefore likely to be a valid index of lexical recruitment; however, future research should examine whether the results observed here extend to other measures of lexical recruitment and other populations (e.g., older adults).

It is also possible that the lack of observed relationship between lexical recruitment and lexically guided perceptual learning is related to a potential threshold effect for a lexical influence on learning. Figure 1 depicts the relationship predicted by both interactive and modular theories in which lexically guided perceptual learning is contingent on lexical recruitment. A potential explanation to reconcile the discrepancy between our findings and these models of speech perception is that only a certain degree of lexical access (or a certain size of the lexicon) is necessary to cue perceptual learning and that beyond this threshold, additional strength in lexical recruitment does not further contribute to lexically guided perceptual learning. That is, lexical contributions could quickly meet a point of diminishing returns. While we observed wide individual variability in the Ganong effect in our two participant samples, it is possible not enough individuals with lexical recruitment at a level below this threshold were recruited, leading to the observed lack of relationship between lexical recruitment and lexically guided perceptual learning in the current work.

This explanation may also contribute to the pattern of results we observed regarding the null effect of language ability on lexically guided perceptual learning. Research on developmental language disorder (and specific language impairment) has suggested that higher-level deficits in receptive language may stem from impairments early in the processing stream, including general auditory processing and global speech perception abilities (e.g., Joanisse & Seidenberg, 2003; McArthur & Bishop, 2004). Despite potential deficits in using bottom-up information to guide speech perception, the current study and previous work (Schwartz et al., 2013) suggest that individuals with lower receptive language ability use top-down lexical information to scaffold speech perception to a higher extent than individuals with higher receptive language ability. It is possible that increased reliance on lexical information is a compensatory mechanism for earlier deficits in speech perception. Compensation of this sort would have benefits not only for online processing, but also for post-perceptual processes. For example, if the relative contribution of lexical information to speech perception is higher in those with weaker receptive language in order to mitigate weaker contributions of phonetic information, then individuals with weaker language ability may surpass the minimal threshold posited above, leading to performance equivalent to those with higher receptive language ability for lexically guided perceptual learning.

Limitations and considerations for future research

Though the current study supports examination of the relationship between receptive language, lexical recruitment, and perceptual learning, the current work does not support the identification of causal mechanisms. While it is plausible that strengthened lexical recruitment in individuals with weaker receptive language ability could be a compensatory mechanism for courser-grained perceptual analysis, the design of the current study does not bear directly on this possibility. Further research is necessary in order to explicate the mechanisms behind the increased reliance on the lexicon observed in both children (Schwartz et al., 2013) and adults with weaker language ability.

As alluded to earlier in the discussion, a problem facing individual differences research in cognitive science more broadly is a lack of knowledge about the degree to which the chosen tasks are stable measures within an individual. While assessments such as the PPVT (used by Colby et al., 2018; Williams & Wang, 1997) and the Trail Making Test (used by Scharenborg et al., 2015; Giovagnoli et al., 1996; Seo et al., 2006) are known to have sufficient test-retest reliability, relatively less is known about stability in performance on speech perception tasks, including the ones used here. That is, there is a dearth of evidence regarding whether performance in the Ganong and/or lexically guided perceptual learning tasks is stable over time – and stimuli – at the level of individual participants. For example, will a person who shows a large Ganong effect for a given stimulus set tested on a Monday also show a large Ganong effect for a different stimulus set when tested on a Friday? As the field advances our efforts to understand individual differences in perceptual and cognitive tasks, additional research is needed in order to confirm that our tasks reflect valid (and thus stable) measures of individual differences.

We acknowledge that this is no mean feat, especially when measuring stability of performance for tasks that assess learning. Recently, Saltzman and Myers (2018) examined whether the size of a perceptual boundary shift induced by lexically guided perceptual learning was consistent in individuals who completed the same lexically guided perceptual learning task twice (approximately one week apart). At each session, listeners completed both /s/-bias and /ʃ/-bias exposure phases, and the boundary shift was measured as the difference in /ʃ/-responses between the two phases. They found no relationship in performance across the two sessions, suggesting low individual consistency for lexically guided perceptual learning. However, this study has been retracted (Saltzman & Myers, 2020) and thus it is not clear whether these results are stable. Moreover, assessing the test-retest reliability of this learning paradigm introduces substantial challenges related to disassociating effects of short-term learning from more long-term learning introduced by multiple test sessions. For example, if learning in this paradigm persists over more long-term time periods (as suggested by Eisner & McQueen, 2006; cf. Liu & Jaeger, 2018, 2019), then the *a priori* expectation for individuals who learn would be no correlation between the boundary shift across test sessions because learning from the first session would inherently lead to no learning taking place in a second session. In addition, if an extrinsic factor (as opposed to a stable individual factor) were responsible for a lack of learning in the first session (e.g., completed after a lack of sleep), this may not necessarily lead to no learning occurring in the second session (e.g., completed after a good night’s rest). Thus, it is impossible to completely rule out insufficient stability of this effect as a contributor to null results.

Additional research regarding the stability of both of these effects within an individual is warranted not only to explicate the theoretical relationship between lexical recruitment and lexically guided perceptual learning, but also to support clinical use of these tasks. Results from past research (Schwartz et al., 2013) and the current study have shown that a larger Ganong effect is associated with SLI and weaker receptive language ability, respectively. Previous research has demonstrated that a larger Ganong effect is also associated with weaker speech in noise perception (Lam et al., 2017). Given that weaker speech in noise perception has been shown to be predictive of language impairment (Ziegler et al., 2005, 2011) and that language impairment is associated with broad receptive language deficits, it is possible that once a better understanding of factors contributing to the individual differences and internal consistency of the Ganong effect is gained, this task could become a valuable, time-effective tool for use in clinical batteries for the assessment of language impairment.

Conclusions

The findings of Experiment 1 are consistent with a theory positing that individuals with weaker language ability demonstrate increased reliance on lexical information for speech perception compared to those with stronger receptive language ability. Increased reliance on lexical information among those with weaker receptive language ability was observed for online lexical recruitment, but no differences in lexically guided perceptual learning as a function of language ability were observed. Individuals with weaker receptive language ability therefore appear to maintain an intact lexically guided perceptual learning mechanism. Further research is needed in order to understand whether the relationship between lexical recruitment and language ability reflects compensation for earlier deficits in speech perception, and if so, where in the speech processing stream these deficits occur.

To the degree that the chosen measures accurately capture lexical recruitment and lexically guided perceptual learning at the level of individual participants, the findings of both experiments converge to suggest no graded relationship between these two constructs. This result can be accommodated by current theories of speech perception if they are modified to model this relationship as being governed threshold level of lexical recruitment that is necessary and sufficient to cue lexically guided perceptual learning.

Acknowledgments

This work was supported by NIH NIDCD grant R21DC016141 RMT, NSF grants DGE-1747486 and DGE-1144399 to the University of Connecticut, and by the Jorgensen Fellowship (University of Connecticut) to NG. The views expressed here reflect those of the authors and not the NIH, the NIDCD, or the NSF. Portions of this study were presented at the 177th meeting of the Acoustical Society of America. We extend gratitude to Emily Myers for providing her stimuli for use in this study; gratitude is also extended to Lee Drown, Amanda Salemi, Emma Hungaski, and Andrew Pine for assistance with administration and scoring of the assessment battery.

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*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International*, *5*.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence: TONI-4*. Pro-Ed.

Brown, V. A., Hedayati, M., Zanger, A., Mayn, S., Ray, L., Dillman-Hasso, N., & Strand, J. F. (2018). What accounts for individual differences in susceptibility to the McGurk effect? *PloS One*, *13*(11).

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.

Colby, S., Clayards, M., & Baum, S. (2018). The role of lexical status and individual differences for perceptual learning in younger and older adults. *Journal of Speech, Language, and Hearing Research*, *61*(8), 1855–1874.

Drouin, J. R., & Theodore, R. M. (2018). Lexically guided perceptual learning is robust to task-based changes in listening strategy. *The Journal of the Acoustical Society of America*, *144*(2), 1089–1099.

Drouin, J. R., Theodore, R. M., & Myers, E. B. (2016). Lexically guided perceptual tuning of internal phonetic category structure. *The Journal of the Acoustical Society of America*, *140*(4), EL307–EL313.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950–1953.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125.

Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacona, M., & Capitani, E. (1996). Trail making test: Normative values from 287 normal adult controls. *The Italian Journal of Neurological Sciences*, *17*(4), 305–309.

Hsu, H. J., & Bishop, D. V. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science*, *17*(3), 352–365.

Ishida, M., Samuel, A. G., & Arai, T. (2016). Some people are “more lexical” than others. *Cognition*, *151*, 68–75.

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language*, *86*(1), 40–56.

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.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–178.

Lam, B. P., Xie, Z., Tessmer, R., & Chandrasekaran, B. (2017). The downside of greater lexical influences: Selectively poorer speech perception in noise. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1662–1673.

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*(5), 358.

Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition*, *174*, 55–70.

Liu, L., & Jaeger, T. F. (2019). Talker-specific pronunciation or speech error? Discounting (or not) atypical pronunciations during speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(12), 1562.

McArthur, G. M., & Bishop, D. V. M. (2004). Frequency discrimination deficits in people with specific language impairment. *Journal of Speech, Language, and Hearing Research*.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, *50*(1–3), 271–285.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review*, *13*(6), 958–965.

Myers, E. B., & Mesite, L. M. (2014). Neural systems underlying perceptual adjustment to non-standard speech tokens. *Journal of Memory and Language*, *76*, 80–93.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–325.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 1037–1052.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*(3), 271–276.

Saltzman, D., & Myers, E. (2018). Listeners are maximally flexible in updating phonetic beliefs over time. *Psychonomic Bulletin & Review*, *25*(2), 718–724.

Saltzman, D., & Myers, E. (2020). Retraction Note: Listeners are maximally flexible in updating phonetic beliefs over time. *Psychonomic Bulletin & Review*, *27*, 819. https://doi.org/10.3758/s13423-020-01765-0

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207–1218. https://doi.org/10.3758/APP.71.6.1207

Scharenborg, O., Weber, A., & Janse, E. (2015). The role of attentional abilities in lexically guided perceptual learning by older listeners. *Attention, Perception, & Psychophysics*, *77*(2), 493–507.

Schwartz, R. G., Scheffler, F. L., & Lopez, K. (2013). Speech perception and lexical effects in specific language impairment. *Clinical Linguistics & Phonetics*, *27*(5), 339–354.

Seo, E. H., Lee, D. Y., Kim, K. W., Lee, J. H., Jhoo, J. H., Youn, J. C., Choo, I. H., Ha, J., & Woo, J. I. (2006). A normative study of the Trail Making Test in Korean elders. *International Journal of Geriatric Psychiatry*, *21*(9), 844–852.

Theodore, R. M., Monto, N. R., & Graham, S. (2019). Individual differences in distributional learning for speech: What's ideal for ideal observers? *Journal of Speech, Language, and Hearing Research*, 1–13.

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399–433.

Wiig, E. H., Semel, E., & Secord, W. (2013). Clinical Evaluation of Language Fundamentals—Fifth Edition. *Bloomington, MN: Pearson*.

Williams, K., & Wang, J. (1997). *Technical References to the Peabody Picture Vocabulary Test-(PPVT-III)*. American Guidance Services.

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072.

Xie, X., Theodore, R. M., & Myers, E. B. (2017). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 206–217. https://doi.org/10.1037/xhp0000285

Zheng, Y., & Samuel, A. G. (2020). The relationship between phonemic category boundary changes and perceptual adjustments to natural accents. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(7), 1270–1292. https://doi.org/10.1037/xlm0000788

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences*, *102*(39), 14110–14115.

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. *Journal of Experimental Child Psychology*, *110*(3), 362–372.

Table 1. Results of the four mixed effects models for /k/ responses in the Ganong task for Experiment 1 that included fixed effects of VOT, continuum, and CELF subtest (each in a separate model).

Model	**Fixed effect**	$\hat{\beta}$	***SE***	***z***	***p***
Formulated Sentences (FS)	(Intercept)	-1.164	0.138	-8.433	< 0.001
	VOT	3.539	0.156	22.666	< 0.001
	Continuum	1.078	0.196	5.508	< 0.001
	FS	-0.055	0.136	-0.402	0.688
	VOT x Continuum	1.149	0.221	5.191	< 0.001
	VOT x FS	0.333	0.150	2.220	0.026
	Continuum x FS	**-0.021**	**0.189**	**-0.112**	**0.911**
	VOT x Continuum x FS	0.013	0.201	0.064	0.949
Recalling Sentences (RS)	(Intercept)	-1.022	0.122	-8.385	< 0.001
	VOT	3.605	0.164	22.006	< 0.001
	Continuum	1.351	0.188	7.190	< 0.001
	RS	0.032	0.120	0.266	0.790
	VOT x Continuum	1.128	0.217	5.200	< 0.001
	VOT x RS	0.428	0.159	2.690	0.007
	Continuum x RS	**-0.422**	**0.182**	**-2.318**	**0.020**
	VOT x Continuum x RS	-0.130	0.198	-0.659	0.510
Understanding Spoken Paragraphs (USP)	(Intercept)	-1.076	0.113	-9.503	< 0.001
	VOT	3.463	0.150	23.035	< 0.001
	Continuum	1.266	0.169	7.487	< 0.001
	USP	-0.060	0.111	-0.541	0.589
	VOT x Continuum	1.104	0.187	5.904	< 0.001
	VOT x USP	0.204	0.142	1.434	0.151
	Continuum x USP	**-0.434**	**0.163**	**-2.667**	**0.008**
	VOT x Continuum x USP	-0.025	0.156	-0.162	0.871
Semantic Relationships (SR)	(Intercept)	-1.043	0.111	-9.408	< 0.001
	VOT	3.469	0.161	21.572	< 0.001
	Continuum	1.158	0.175	6.627	< 0.001
	SR	-0.176	0.109	-1.618	0.106
	VOT x Continuum	1.098	0.188	5.836	< 0.001
	VOT x SR	0.183	0.155	1.182	0.237
	Continuum x SR	**-0.459**	**0.169**	**-2.711**	**0.007**
	VOT x Continuum x SR	0.200	0.160	1.253	0.210

Table 2. Single-order correlations between task performance and CELF subtest standard scores in Experiment 1. As described in the main text, Ganong performance was quantified as the difference in proportion /k/ responses between the *giss-kiss* and *gift-kift* continua. Lexically guided perceptual learning performance was quantified as the difference in proportion /s/ responses between the first test (following /s/-bias exposure) and the second test (following /ʃ/-bias exposure.

Task	**CELF subtest**	***r***	***p***
Ganong	Formulated Sentences	-0.09	0.502
	Recalling Sentences	-0.37	0.004
	Understanding Spoken Paragraphs	-0.35	0.003
	Semantic Relationships	-0.27	0.031
Lexically guided perceptual learning	Formulated Sentences	-0.07	0.622
	Recalling Sentences	0.02	0.856
	Understanding Spoken Paragraphs	-0.07	0.546
	Semantic Relationships	0.05	0.700

Table 3. Results of the four mixed effects models for /s/ responses in the lexically guided perceptual learning task for Experiment 1 that included fixed effects of step, bias, and CELF subtest (each in a separate model).

Model	**Fixed effect**	$\hat{\beta}$	***SE***	***z***	***p***
Formulated Sentences (FS)	(Intercept)	3.011	0.380	7.915	< 0.001
	Step	4.108	0.265	15.505	< 0.001
	Bias	1.461	0.508	2.874	< 0.001
	FS	0.158	0.345	0.458	0.647
	Step x Bias	0.580	0.487	1.192	0.233
	Step x FS	0.310	0.205	1.513	0.130
	Bias x FS	**-0.259**	**0.385**	**-0.674**	**0.500**
	Step x Bias x FS	-0.082	0.330	-0.248	0.804
Recalling Sentences (RS)	(Intercept)	2.860	0.348	8.223	< 0.001
	Step	4.026	0.241	16.716	< 0.001
	Bias	1.401	0.438	3.197	0.001
	RS	-0.067	0.314	-0.212	0.832
	Step x Bias	0.420	0.441	0.950	0.342
	Step x RS	0.318	0.194	1.636	0.102
	Bias x RS	**-0.249**	**0.314**	**-0.793**	**0.428**
	Step x Bias x RS	-0.269	0.313	-0.858	0.391
Understanding Spoken Paragraphs (USP)	(Intercept)	2.880	0.322	8.947	< 0.001
	Step	3.981	0.227	17.568	< 0.001
	Bias	1.484	0.406	3.652	< 0.001
	USP	0.059	0.294	0.200	0.841
	Step x Bias	0.440	0.393	1.120	0.263
	Step x USP	-0.014	0.184	-0.073	0.942
	Bias x USP	**-0.500**	**0.302**	**-1.655**	**0.098**
	Step x Bias x USP	-0.389	0.273	-1.424	0.154
Semantic Relationships (SR)	(Intercept)	2.894	0.342	8.451	< 0.001
	Step	4.087	0.246	16.604	< 0.001
	Bias	1.575	0.447	3.523	< 0.001
	SR	-0.084	0.310	-0.271	0.787
	Step x Bias	0.580	0.432	1.343	0.179
	Step x SR	0.012	0.196	0.062	0.951
	Bias x SR	**-0.383**	**0.327**	**-1.171**	**0.241**
	Step x Bias x SR	-0.609	0.291	-2.094	0.036

Figure captions

Figure 1. Illustration of the process by which lexical information leads to lexically guided perceptual learning according to interactive and modular models of speech perception.

Figure 2. Beeswarm plots showing individual variation of standard scores for the four CELF subtests administered in Experiment 1. Expressive language measures are shown in blue; receptive language measures are shown in gray. Points are jittered along the x-axis to promote visualization of overlapping scores.

*Figure 3***.** Mean proportion /k/ responses at each VOT for each continuum in the control and Ganong blocks for the preliminary experiment. Means reflect grand means calculated over by-subject averages. As described in the main text, stimuli presented in the control block contained the initial CV portion of stimuli presented in the Ganong block. Error bars indicate standard error of the mean.

*Figure 4***.** Mean proportion /k/ responses at each VOT for each continuum in the Ganong task. Experiment 1 is shown in panel A; Experiment 2 is shown in panel B. Means reflect grand means calculated over by-subject averages. Error bars indicate standard error of the mean.

*Figure 5***.** Panel A displays the beta estimate for the subtest by continuum interaction in each of the mixed effects models shown in Table 1; error bars indicate the 95% confidence interval. The expressive language measures are shown in blue and the receptive language measures are shown

in gray. Panel B shows proportion /k/ responses at each VOT for each continuum according to a median split of participants by Understanding Spoken Paragraphs (USP) score; error bars indicate standard error of the mean. As described in the main text, subtest score was entered as a continuous variable in all models. The median split displayed here is to illustrate the nature of the subtest by continuum interaction for the receptive language measures.

Figure 6. Mean proportion /s/ responses following each bias exposure block for each step of the test continuum in the perceptual learning task. Experiment 1 is shown in panel A; Experiment 2 is shown in panel B. Facets separate performance into the first and second halves of each test block. Error bars indicate standard error of the mean.

*Figure 7***.** Panel A displays the beta estimate for the subtest by bias interaction in each of the mixed effects models shown in Table 3; error bars indicate the 95% confidence interval. The expressive language measures are shown in blue and the receptive language measures are shown in gray. Panel B shows proportion /s/ responses at each step for each bias condition according to a median split of participants by Understanding Spoken Paragraphs (USP) score; error bars indicate standard error of the mean. As described in the main text, subtest score was entered as a continuous variable in all models. The median split displayed here is to illustrate the nature of the subtest by continuum interaction for the receptive language measures.

Figure 8. Scatterplot illustrating the null relationship between the magnitude of the lexically guided perceptual learning effect (LGPL) and the Ganong effect in Experiment 1 (panel A) and

Experiment 2 (panel B). As described in the main text, higher values indicate larger effects along both axes.