Title
Chunking Ability Shapes Sentence Processing at Multiple Levels of Abstraction

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
McCauley, Stewart M.
Isbilen, Erin S.
Christiansen, Morten H.

Publication Date
2017

Peer reviewed

Chunking Ability Shapes Sentence Processing at Multiple Levels of Abstraction

Stewart M. McCauley (Stewart.McCauley@liverpool.ac.uk)
Department of Psychological Sciences, University of Liverpool

Erin S. Isbilen (esi6@cornell.edu)
Morten H. Christiansen (christiansen@cornell.edu)
Department of Psychology, Cornell University

Abstract

Several recent empirical findings have reinforced the notion that a basic learning and memory skill—*chunking*—plays a fundamental role in language processing. Here, we provide evidence that chunking shapes sentence processing at multiple levels of linguistic abstraction, consistent with a recent theoretical proposal by Christiansen and Chater (2016). Individual differences in chunking ability at two different levels is shown to predict on-line sentence processing in separate ways: i) phonological chunking ability, as assessed by a variation on the non-word repetition task, predicts processing of complex sentences featuring phonological overlap; ii) multiword chunking ability, as assessed by a variation on the serial recall task, is shown to predict reading times for sentences featuring long-distance number agreement with locally distracting number-marked nouns. Together, our findings suggest that individual differences in chunking ability shape language processing at multiple levels of abstraction, consistent with the notion of language acquisition as learning to process.

Keywords: sentence processing; chunking; learning; memory; usage-based approach; language

Introduction

Language takes place in real time; a fairly uncontroversial observation, yet one with far-reaching consequences that are rarely considered. For instance, a typical English speaker produces between 10 and 15 phonemes per second (Studdert-Kennedy, 1986), yet the ability of the auditory system to process discrete sounds is limited to around 10 per second, beyond which the signal is perceived as a single buzz (Miller & Taylor, 1948). Moreover, the auditory trace is limited to about 100ms (Remez et al., 2010). Compounding matters even further, human memory for sequences is limited to between 4 and 7 items (e.g., Cowan, 2001; Miller, 1956). Simply put, the sensory signal is so incredibly short-lived, and our memory for it so very limited, that language would seem to stretch the human capacity for information processing beyond its breaking point. We refer to this as the Now-or-Never bottleneck (Christiansen & Chater, 2016).

How is language learning and processing possible in the face of this real-time constraint? A key piece of the puzzle, we suggest, lies in *chunking*: through experience with language, we learn to rapidly recode incoming information into chunks which can then be passed to higher levels of representation.

As an intuitive demonstration of the necessity of chunking, imagine being tasked with recalling a string of letters, presented auditorily: *u o p f m r e e p o a e c s g n p l i r*. After a single presentation of the string, very few listeners would be able to recall a sequence consisting of even half of the letters (cf. Cowan, 2001). However, if exposed to the exact same set of letters but re-ordered slightly, virtually any listener would able to recall the entire sequence with ease: *f r o g m o u s e p a p e r p e n c i l*. Clearly, such a feat is possible by virtue of the ability to rapidly chunk the sequence into familiar sub-sequences (*frog, mouse, paper, pencil*).

According to the proposal of Christiansen and Chater (2016), the Now-or-Never Bottleneck requires language users to perform similar chunking operations on speech and text in order to process and learn from the input. This is necessary both due to the fleeting nature of sensory memory and the speed at which information is encountered during processing. Specifically, language users must perform *Chunk-and-Pass* processing, whereby input is chunked as rapidly as possible and passed to a higher, more abstract level of representation. Information at higher levels must also be chunked before being passed to still higher, increasingly abstract levels of representation.

Thus, in order to communicate in real-time, language users must chunk at multiple levels of abstraction, ranging from the level of the acoustic signal to the level of phonemes or syllables, to words, to multiword units, and beyond. Indeed, mounting empirical evidence supports the notion of chunking at levels higher than that of the individual word: children and adults appear to store and utilize chunks consisting of multiple words in comprehension and production (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008). Moreover, usage-based (e.g., Tomasello, 2003) and generative (e.g., Culicover & Jackendoff, 2005) theoretical approaches have highlighted the importance of such units in grammatical development and sentence processing alike.

Chunking has been considered a key learning and memory mechanism in mainstream psychology for over half a century (e.g., Miller, 1956), and has been used to understand specific aspects of language acquisition (e.g., Jones, 2012; Jones, Gobet, Freudenthal, & Pine, 2014). Nevertheless, few have sought to understand how it may shape more complex linguistic skills, such as sentence processing. McCauley and Christiansen (2015) took an initial step in this direction, showing that individual differences in low-level chunking abilities were predictive of reading times for sentences involving relative clauses, demonstrating the far-reaching impact of basic chunking skills in shaping complex linguistic behaviors.

The present study seeks to evaluate the predictions of the Chunk-and-Pass framework more closely, by examining individual variation in chunking at two different levels of abstraction. Specifically, whereas chunking has previously been treated as a uniform memory ability, we test the novel theoretical prediction that chunking abilities may be relatively independent at different levels of linguistic abstraction. Participants were first asked to take part in a multiword-based serial recall task (Part 1) designed to yield a measure of chunking at the word level. This was followed by a variation on the non-word repetition task (Part 2), designed to yield a measure of phonological chunking ability. Importantly, due to the memory limitations discussed above, participants must utilize chunking in order to recall more than a few discrete words or phonemes in these tasks (e.g., Cowan, 2001; Miller, 1956). Finally, participants took part in an online self-paced reading task (Part 3). The results show that chunking ability at each level predicts different aspects of sentence processing ability: chunking at the phonological level predicts the extent to which low-level phonological information interferes with or facilitates complex sentence processing, while chunking at the multiword level predicts the role of local information in processing sentences with long-distance dependencies.

Part 1: Measuring Individual Differences in Word Chunking Ability

The first task sought to gain a measure of individual participants' ability to chunk words into multiword units. To this end, we specifically isolate chunking as a mechanism by employing a classic psychological paradigm: the serial recall task. Serial recall has a long history of use in studies of chunking, dating back to some of the earliest relevant work (e.g., Miller, 1956), as well being used to extensively study individuals' chunking abilities (e.g., Ericsson, Chase, & Faloon, 1980).

Participants were tasked with recalling strings of 12 individual words, with each string consisting of 4 separate word trigrams extracted from a large corpus of English. Importantly, in order to recall more than a few discrete items (as few as 4 in some accounts; e.g., Cowan, 2001), listeners must chunk the words of the input sequence into larger, multiword units. In this case, we expect them to draw upon linguistic experience with the trigrams in the experimental items.

In addition, we included a baseline performance measure: matched control strings, which featured identical functors to the experimental sequences, along with frequency-matched content words (to avoid semantic overlap effects on recall), presented in random order. Thus, comparing recall for experimental and control trials provides a measure of word chunking ability that reflects language experience while controlling for such factors as attention, motivation, and—to the extent that it is separable—working memory.

Method

Participants 42 native English speakers from the Cornell undergraduate population (17 females; age: *M*=19.8, *SD*=1.2) participated for course credit. Of the original 45 subjects, one was excluded due to audio recording errors, while two subjects failed to complete all three tasks.

Materials Experimental stimuli consisted of word trigrams spanning a range of frequencies, extracted from the American National Corpus (Reppen, Ide & Suderman, 2005) and the Fisher corpus (Cieri, Graff, Kimball, Miller & Walker, 2004). The combined corpus contained a total of 39 million words of American English. Each item was compositional (non-idiomatic). Item frequencies, per million words, ranged from 40 to .08, averaging at .73.

Each word was synthesized independently using the Festival speech synthesizer (Black, Clark, Richmond, King & Zen, 2004) and concatenated into larger strings consisting of 12 words (4 trigrams). Each trigram was matched as closely as possible for frequency with the others occurring in a sequence.

To provide a non-chunk-based control condition, each item was matched to a sequence of words which contained identical functors but random frequency-matched content words (in order to avoid semantic overlap effects on recall, content words were not re-used). The ordering of the words was then randomized. An example of a matched set of sequences is shown below:

1) *have to eat good to know don't like them is really nice*
2) *years got don't to game have she mean to them far is*

The final item set consisted of 20 sequences (10 experimental, 10 control).

Procedure Each trial featured a 12-word sequence presented auditorily. Each word was followed by a 250ms pause. Immediately upon completion of the string, the participant was prompted to verbally recall as much of the sequence as possible. Responses were recorded digitally and later transcribed by a researcher blind to the conditions as well as the purpose of the study.

The presentation order of the sequences was fully randomized. The entire task took approximately 15 minutes.

Results and Discussion

Participants recalled significantly more words from experimental strings than the frequency-matched control sequences. The overall recall rate for words occurring in experimental items was 74.0% (*SE*=2.3%), while the recall rate for control sequences was just 39.2% (*SE*=1.1%). The difference between conditions was significant ($t(41)=18.8$, $p<0.0001$).

As the purpose of Part 1 was to gain an overall measure of chunk sensitivity, we calculated the difference between conditions individually for each subject (*M*=34.8%, *SE*=1.8%), which afforded a measure of word-chunking ability that reflects language experience while controlling for factors such as working memory, attention, and motivation. We refer to this difference measure as the Word Chunk Sensitivity score, and it is used as a predictor of sentence processing ability in Part 3.

In addition to bolstering previous empirical support for compositional (non-idiomatic) multiword sequences as linguistic units in their own right (e.g., Bannard & Matthews, 2008), Part 1 revealed considerable individual differences across participants in word chunking ability. Recall rates for experimental items ranged from as high as 93.3% to just 30.4%, with difference scores across the conditions ranging from 50.8% as low as 3.0%.

Part 2: Measuring Individual Differences in Phonological Chunking Ability

While the first task sought to gain a measure of individual participants' chunking abilities at the level of words, Part 2 sought to gain a measure of chunking ability at the phonological level. To this end, we re-purposed the standard non-word repetition (NWR) task as a *chunking* task. NWR has been used extensively to study various aspects of language development. Recent studies, however, have suggested that chunking may better account for NWR performance than more nebulous psychological constructs, such as working-memory (e.g., Jones, 2012; Jones et al., 2014). In one sense, the NWR task can be re-conceptualized as a serial recall task, as in Part 1. Following such work, and in keeping with the Now-or-Never perspective outlined above, we propose that individual differences in chunking ability underlie differences in NWR performance. In turn, NWR—with appropriately constructed stimuli—can serve as an additional dimension along which to measure chunking ability at the level of phonological processing.

Participants engaged in a standard NWR task, with each non-word consisting of 4, 5, or 6 syllables. However, the stimuli were designed such that the same set of syllables occurred in two different non-words, but in different orderings: one ordering yielded an item with high "chunkability," according to corpus statistics, while the other was estimated to be less "chunkable." The two items were then counterbalanced across halves of the task.

Method

Participants The same 42 subjects from Part 1 participated directly afterwards in this task.

Materials Non-words were generated using an algorithm which took a large list[1] of English syllables and randomly generated syllable combinations that were evaluated according to distributional statistics at the phoneme level. For the purpose of supplying statistics, the combined corpus used in Part 1 was automatically re-transcribed phonetically using the Festival speech synthesizer (Black et al., 2004).

For each of three different syllable lengths (4-, 5-, and 6-syllables), the algorithm extracted item pairs that differed maximally in sequence likelihood (based on phoneme trigram statistics) across two different sequential orderings of the same set of syllables. In other words, pairs were selected in which one ordering of syllables was highly "chunk-like," while the other ordering of the same syllables was less "chunk-like," according to the phoneme statistics of the corpus. Four sets of non-words (the four in which the pair differed most greatly in terms of sequence likelihood) were selected for each syllable length. An example of a highly "chunk-like" 4-syllable item is *krew-ih-tie-zuh*, which was matched to the less chunk-like *tie-zuh-ih-krew*.

Thus, the final set of items included 24 non-words, eight in each of three syllable-length conditions (4-, 5-, and 6-syllable), with four being highly "chunk-like" and the other four consisting of alternate orderings of the same syllables which were statistically less "chunk-like."

Procedure The task was split into two blocks, with all NWR item pairs counterbalanced between them. The auditory presentation of each non-word was followed by a 1500ms pause, after which the participant was prompted to recall the item verbally. As with Part 1, responses were recorded digitally and scored offline. The task took approximately 4 minutes to complete.

Correct responses received a score of 1. Responses involving alteration to a single phoneme (usually a vowel substitution, which could easily stem from differences in regional dialect) received a score of 0.5. All other responses received scores of 0.

Results and Discussion

Participants achieved a mean NWR accuracy rate of 54.1% (*SE*=2.3%). While the overall differences between the high chunk-like (M=55.2%, SE=2.5%) and low chunk-like (M=53.1%, SE=2.5%) conditions were in the expected direction, they were subtle, with a mean difference of 2.1% (non-significant: $t(41)=1.12$, $p>0.1$). However, there was considerable individual variation in the size of this difference across participants (*SE*=1.9%), ranging from 29.2% to less than 0%, at -16.6%. Therefore, in Part 3, we assess both the overall NWR performance score as well as the difference between the conditions (which we refer to as the Phonological Chunk Sensitivity score) as predictors of sentence processing.

Importantly, neither the overall raw task performance (β=-0.03, p=0.9) nor the Chunk Sensitivity scores (β=-0.19, p=0.22) from Parts 1 and 2 correlated with one another, consistent with the notion that chunking at each level may have different consequences for sentence processing.

Part 3: Measuring Individual Differences in Sentence Processing and Chunking

In Part 1, we sought to gain a measure of individual participants' ability to chunk words together, while Part 2 aimed to provide a measure of phonological chunking ability. In Part 3, the same subjects from the first two parts

[1] http://semarch.linguistics.fas.nyu.edu/barker/Syllables/

participated in a self-paced reading task designed to: i) assess on-line sentence processing across two different sentence types which were hypothesized to involve chunking at the word and phonological levels, but to different extents; ii) determine the extent to which chunking ability, as assessed in the first two tasks, predicted processing difficulties for each sentence type.

The first sentence type featured long distance subject-verb number agreement with locally distracting number-marked nouns, exemplified by (1):

1. *The **key** to the <u>cabinets **was**</u> rusty from many years of disuse.*

Previous work (Pearlmutter, Garnsey, & Bock, 1999) has shown that readers are slower to process the verb when the number of the local noun (*cabinets*) does not match that of the head noun (*key*), resulting in the sequence (*cabinets was*). Reading times are compared to sentences in which the number marking matches, as exemplified by (2):

2. *The **key** to the <u>cabinet **was**</u> rusty from many years of disuse.*

In other words, reading times are higher at the verb when the local information is distracting. Following the finding that text-chunking ability predicts decreased difficulty with complex sentences involving long-distance dependencies (McCauley & Christiansen, 2015), we hypothesized that participants with higher Word Chunk Sensitivity scores (Part 1) would be less susceptible to interference from local information in sentences such as (1). Subjects that are better able to rapidly chunk words together and pass them to higher levels of representation should not only experience decreased computational burden from long-distance dependencies, but should be less affected by locally distracting information.

The second sentence type featured object-relative (OR) clauses, which have been shown to be processed with greater ease by good text chunkers (McCauley & Christiansen, 2015). However, in the present study we added an element of phonological interference: two pairs of words in each sentence exhibited phonological overlap. Previous work has shown that low-level phonological overlap can interfere with the processing of sentences featuring relative clauses (Acheson & MacDonald, 2011). An experimental item and its matched control are shown in (3) and (4):

3. *The <u>cook</u> that the <u>crook</u> **consoles controls** the politician.*
4. *The prince that the crook comforts controls the politician*.

In line with the Chunk-and-Pass framework, we predicted that better phonological chunkers, as assessed in Part 2, would be less susceptible to phonological interference, by virtue of their ability to more rapidly chunk and pass phonological information to a higher level of representation.

Thus, participants' resilience to phonological interference was hypothesized to be better predicted by Phonological Chunk Sensitivity (Part 2), while participants' susceptibility to local number mismatch was expected to be better predicted by Word Chunk Sensitivity (Part 1).

Method

Participants The same 42 subjects from Parts 1 and 2 participated in Part 3 immediately afterwards.

Materials There were two sentence lists—counterbalanced across subjects—each consisting of 9 practice items, 20 experimental items, 20 matched control items, and 68 filler items. There were two experimental conditions, each with 20 items; the first consisted of the OR sentences featuring phonological overlap (the first 20 items from Acheson & MacDonald, 2011). The second experimental condition consisted of grammatical sentences featuring long-distance number agreement with locally distracting number-marked nouns (the 16 items from Pearlmutter et al., 1999, plus four additional sentences with the same properties).

Each list included, for each condition, 10 of the items in their experimental form and 10 of the items in their control form (without rhymes in the case of the OR sentences; without locally distracting nouns in the case of the number agreement sentences). The lists were counterbalanced such half of the subjects saw the experimental versions of sentences the other half saw in their control form.

Procedure Materials were presented in random order using a self-paced, word-by-word moving window display (Just, Carpenter, & Woolley, 1982). At the beginning of each trial, a series of dashes appeared (one corresponding to each nonspace character in the sentence). The first press of a marked button caused the first word to appear, while subsequent button presses caused each following word to appear. The previous word would return once more to dashes. Reaction times were recorded for each button press. Following each sentence, subjects answered a yes/no comprehension question using buttons marked "Y" and "N." The task took approximately 10 minutes.

Results and Discussion

Only trials with correct answers to comprehension questions were analyzed. Accuracy for the number agreement condition was 88.3%; for the object-relatives it was 80.0%. Following Acheson & MacDonald (2011), raw reaction times over 3000ms were excluded. Prior to analysis, raw reaction times (RTs) were log-transformed.

Mean RTs for the main verb in the number agreement and phonological overlap sentences were comparable to those in the corresponding original studies (respectively: Pearlmutter et al., 1999; Acheson & MacDonald, 2011), as was the size of the mean difference between conditions. In the number agreement condition, the verb in experimental items (M=361.1, SE=19.9) was processed more slowly than in controls (M=316.7, SE=13.9), a mean difference of 44ms ($F1[1,41]$=12.7, p<0.001; $F2[1,18]$=10.2, p<0.01). There was a fair amount of individual variation in the difference

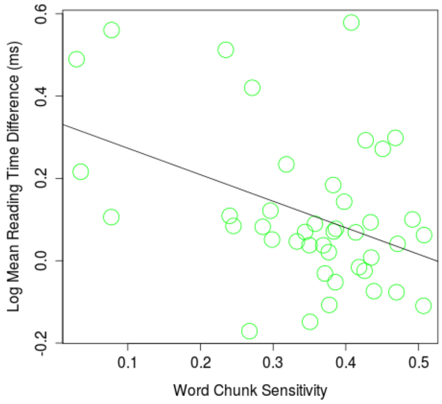

Fig. 1: Correlation between Word Chunk Sensitivity (derived from recall scores in Part 1) and the difference in main verb RTs between sentences with locally distracting number information vs. control sentences.

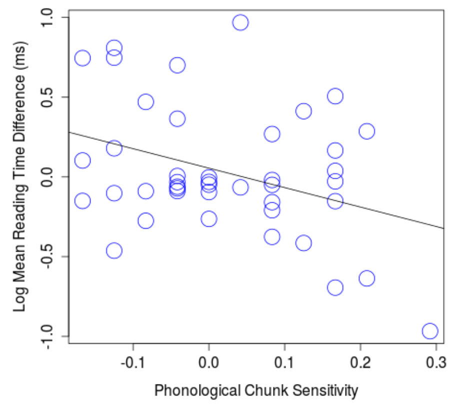

Fig. 2: Correlation between Phonological Chunk Sensitivity (derived from repetition scores in Part 2) and the difference in main verb RTs for OR sentences with and without phonological overlap between words.

between conditions (SD=79.4).

The critical main verb in OR sentences featuring phonological overlap was processed more slowly (M=605.1, SE=70.6) than in matched controls (M=546.3, SE=42.2), a mean difference of 58.8 which was non-significant (*F1[1,41]*=1.21, *p*=0.28; *F2[1,18]*=0.04, *p*=0.8; see discussion). There was, however, considerable individual variation in the difference between conditions (SD=343.7), especially relative to the size of group mean difference.

We were primarily interested in the extent to which differences in RTs between experimental and control sentences could be predicted by the Chunk Sensitivity measures collected in Parts 1 and 2. Below, we analyze these relationships using multiple linear regression, with Word Chunk Sensitivity and Phonological Chunk Sensitivity scores as predictors of RT differences between conditions (recall that the two metrics were not correlated).[2]

For the difference between sentences featuring locally distracting number information and their control counterparts, we found that Word Chunk Sensitivity was a significant predictor of RT difference at the verb (β=-0.79, *t*=-3.19, *p*<0.01), while Phonological Chunk Sensitivity and the interaction term did not reach significance. The model for the significant main effect had an *R* value of 0.42. The correlation between Word Chunk Sensitivity and the RT difference is depicted in Figure 1. As can be seen, subjects with higher Word Chunk Sensitivity scores appear less susceptible to interference from the locally distracting number information, as reflected by lower differences between verb RTs for experimental vs. control sentences.

With regard to the difference between OR sentences with and without phonological overlap, we found that Phonological Chunk Sensitivity was a significant predictor of RT differences at the main verb (β=-3.49, *t*=-2.43, *p*<0.05), while Word Chunk Sensitivity and the interaction term did not reach significance. The model for the significant main effect had an *R* value of 0.36. A scatterplot showing the correlation between Phonological Chunk Sensitivity and the RT difference is shown in Figure 2: better chunking ability resulted in less phonological interference.

Thus, consistent with the predictions of the Chunk-and-Pass framework, we find evidence for the notion that chunking ability shapes sentence processing differently at two separate levels of abstraction: participants who were more sensitive to word chunk information better processed long-distance dependencies in the face of conflicting local information, while those with higher phonological chunk sensitivity better processed complex sentences with phonological overlap between words. That the two chunk sensitivity measures did not correlate with one another further underscores the notion of chunking taking place at multiple levels of abstraction.

While we failed to find the same effect of phonological overlap on processing as did Acheson and MacDonald (2011), it is likely that our subjects (Cornell undergraduates) had more reading experience than subjects at UW-Madison, and experienced less interference overall. Nonetheless, our measure of phonological chunk sensitivity was sensitive enough to pick up individual differences that predicted sentence processing in the face of phonological interference.

Intriguingly, participants with very high Phonological Chunk Sensitivity appeared to experience an *advantage* for OR sentences featuring phonological overlap. This raises the possibility that such subjects benefitted from phonologically-based priming of subsequent rhyme words in sentences such as (3). Further work will be necessary to evaluate this possibility.

General Discussion

In the present study, we show that individual differences in chunking ability predict on-line sentence processing at multiple levels of abstraction: chunking at the phonological level is shown to predict the way phonological information

[2] We found that raw NWR performance scores resulted in weaker linear models and did not reach significance as a predictor. Therefore, we focus on the Phonological Chunk Sensitivity metric in the analyses (see Part 2).

is used during complex sentence processing, while chunking at the multiword level is shown to predict the ease with which long-distance dependencies are processed in the face of conflicting local syntactic information. In Part 1, we adapted the serial recall task—a paradigm used for over half a century to study memory, including chunking phenomena—in order to gain a measure of individual variation in subjects' ability to chunk word sequences into multiword units. In Part 2, subjects participated in a NWR task with non-words designed to vary according to the ease with which their phonemes could be chunked. The difference in correct repetition rates between highly chunk-able and less chunk-able items provided a measure of individual variation in chunking ability at the phonological level. Finally, in Part 3 we showed that chunking at the multiword level was predictive of processing for sentences with long-distance dependencies and distracting local information, while chunking at the phonological level was predictive of complex sentence processing in the presence of phonological overlap between words.

Expanding on the findings of a previous study that showed low-level chunking of sub-lexical letter sequences to predict sentence processing abilities (McCauley & Christiansen, 2015), the present study supports the notion that chunking not only takes place at multiple levels of abstraction, but that individuals' processing abilities may be differently shaped by chunking at each level. Moreover, chunking at lower levels (e.g., the phonological level) may have serious consequences for processing at higher levels (e.g., sentence processing).

This work is highly relevant to the study of language acquisition. The Now-or-Never bottleneck imposes incremental, on-line processing constraints on language learning, suggesting a key role for chunking. Indeed, a number of recent computational modeling studies have demonstrated that chunking can account for key empirical findings on children's phonological development and word learning abilities (Jones, 2012; Jones et al., 2014), while other work has captured a role for chunking in learning to comprehend and produce sentences (McCauley & Christiansen, 2011, 2014). There exists a clear need for further developmental behavioral studies—including longitudinal studies—examining individual differences in chunking as they pertain to specific stages of language development as well as more general language learning outcomes.

Acknowledgments

Thanks to Nick Chater and Gary Jones for helpful discussion, as well as S. Reig, K. Diamond, J. Kolenda, J. Powell, S. Goldberg, and D. Dahabreh for assistance with participant running and recruitment.

References

Acheson, D.J. & MacDonald, M.C. (2011). The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. *Journal of Memory and Language, 65,* 193-207.

Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62,* 67-82.

Bannard, C. & Matthews, D. (2008). Stored word sequences in language learning. *Psychological Science, 19,* 241-248.

Christiansen, M.H. & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences, 39,* e62.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87-114.

Culicover, P.W. & Jackendoff, R. (2005). *Simpler syntax.* New York: Oxford University Press.

Ericsson, K.A., Chase, W.G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208,* 1181-1182.

Jones, G. (2012). Why chunking should be considered as an explanation for developmental change before short-term memory capacity and processing speed. *Frontiers in Psychology, 3 :167*. DOI: 10.3389/fpsyg.2012.00167.

Jones, G., Gobet, F., Freudenthal, D., Watson, S.E. & Pine, J.M. (2014). Why computational models are better than verbal theories: The case of nonword repetition. *Developmental Science, 17,* 298-310.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228-238.

McCauley, S.M. & Christiansen, M.H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1619-1624). Austin, TX: Cognitive Science Society.

McCauley, S.M. & Christiansen, M.H. (2014). Acquiring formulaic language: A computational model. *Mental Lexicon, 9,* 419-436.

McCauley, S.M. & Christiansen, M.H. (2015). Individual differences in chunking ability predict on-line sentence processing. In D.C. Noelle & R. Dale (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Miller, G.A. & Taylor, W.G. (1948). The perception of repeated bursts of noise. *Journal of the Acoustic Society of America, 20*, 171-182.

Pearlmutter, N.J., Garnsey, S.M. & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41,* 427-456.

Remez, R.E., Ferro, D.F., Dubowski, K.R., Meer, J., Broder, R.S. & Davids, M.L. (2010). Is desynchrony tolerance adaptable in the perceptual organization of speech? *Attention, Perception, & Psychophysics, 72,* 2054-2058.

Studdert-Kennedy, M. (1986). Some developments in research on language behavior. In N.J. Smelser & D.R. Gerstein (Eds.), *Behavioral and social science: Fifty years of discovery* (pp. 208-248). Washington, DC: National Academy Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.