Why computational models are better than verbal theories: The case of nonword repetition

Gary Jones[1], Fernand Gobet[2], Daniel Freudenthal[3], Sarah E. Watson[1] and Julian M. Pine[3]

[1]Nottingham Trent University, [2]Brunel University, [3]University of Liverpool (all UK)

Author Note

Correspondence concerning this article should be addressed to Gary Jones, Division of Psychology, Nottingham Trent University, Chaucer Building, Nottingham, NG1 5LT (UK). E-mail: gary.jones@ntu.ac.uk.

Abstract

Tests of nonword repetition (NWR) have often been used to examine children's phonological knowledge and word learning abilities. However, theories of NWR primarily explain performance either in terms of phonological working memory or long-term knowledge, with little consideration of how these processes interact. One theoretical account that focuses specifically on the interaction between short-term and long-term memory is the chunking hypothesis. Chunking occurs because of repeated exposure to meaningful stimulus items, resulting in the items becoming grouped (or chunked); once chunked, the items can be represented in short-term memory using one chunk rather than one chunk per item. We tested several predictions of the chunking hypothesis by presenting 5-6 year-old children with three tests of NWR that were either high, medium, or low in wordlikeness. The results did not show strong support for the chunking hypothesis, suggesting that chunking fails to fully explain children's NWR behavior. However, simulations using a computational implementation of chunking (namely CLASSIC, or Chunking Lexical And Sub-lexical Sequences In Children) show that, when the linguistic input to 5-6 year old children is estimated in a reasonable way, the children's data is matched across all three NWR tests. These results have three implications for the field: (a) a chunking account can explain key NWR phenomena in 5-6 year old children; (b) tests of chunking accounts require a detailed specification both of the chunking mechanism itself and of the input on which the chunking mechanism operates; and (c) verbal theories emphasizing the role of long-term knowledge (such as chunking) are not precise enough to make detailed predictions about experimental data, but computational implementations of the theories can bridge the gap.

Introduction

Nonword repetition (NWR) is a relatively simple developmental task in which children repeat aloud a set of nonwords spoken to them. Strong associations exist between NWR performance and language ability. For example, NWR performance at four years of age is a strong predictor of vocabulary size at five years of age (Gathercole & Baddeley, 1989); NWR is a significant predictor of children's ability to learn a second language (e.g., Cheung, 1996; Masoura & Gathercole, 2005; Service, 1992); and NWR is a key marker of language impairment (e.g., Bishop, North, & Donlan, 1996; Conti-Ramsden, Botting, & Faragher, 2001; Weismer, Tomblin, Zhang, Buckwalter, Chynoweth, & Jones, 2000).

Research involving NWR has suggested that there are both short-term memory and long-term memory components to the task. On the one hand, long nonwords are repeated less accurately than short nonwords, suggesting that short-term memory capacity may be involved (e.g., Weismer et al., 2000; Gathercole & Baddeley, 1989; Stokes, Wong, Fletcher, & Leonard, 2006). On the other hand, repetition accuracy is greater for nonwords that are rated as wordlike compared to nonwords that are not rated as wordlike (e.g., Gathercole, 1995; Munson, Kurtz, & Windsor, 2005); and nonwords containing phoneme sequences that occur frequently in the native language are repeated more accurately than nonwords containing phoneme sequences that occur infrequently in the native language (e.g., Edwards, Beckman, & Munson, 2004; Vitevich, Luce, Charles-Luce, & Kemmerer, 1997). The latter two findings suggest that long-term memory is also involved in NWR.

Prior to detailing theoretical explanations of NWR performance, some terminology is needed to specify the use of lexical and sub-lexical knowledge in

NWR. Henceforth, we use the term *wordlikeness* to refer to the degree to which a nonword contains lexical and sub-lexical units (these can be lexical or morphological e.g., ru<u>bid</u>, glister<u>ing</u>; or they can be phoneme sequences that frequently occur, e.g., d<u>ake</u>, g<u>uck</u>); *wordlikeness effect* to describe any influence that lexical and sub-lexical knowledge has on repetition performance; and *phonological knowledge* to refer to long-term knowledge of phonemes and phoneme sequences.

Theoretical explanations of nonword repetition

In line with the pattern of NWR performance outlined above, theoretical explanations tend to favor either short-term working memory or long-term knowledge. Gathercole and colleagues (e.g., Baddeley, Gathercole, & Papagno, 1998; Gathercole & Baddeley, 1989; Gathercole, 2006) argue that phonological working memory plays a primary role in NWR performance. Nonwords are stored in phonological working memory before being repeated aloud. Because long nonwords occupy more space in working memory, the representation of long nonwords is more likely to be compromised than the representation of short nonwords (Baddeley, 1986; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002).

Alternative explanations of NWR focus on long-term linguistic knowledge. For example, Metsala (1999) suggests that NWR performance is influenced by lexical restructuring. As children learn new words, there is a drive to further elaborate the phonological knowledge of words from dense rather than sparse neighborhoods in order to differentiate their similar sounds. In a similar vein, Munson and colleagues (Munson, Edwards, & Beckman, 2005; Munson, Kurtz, & Windsor, 2005) suggest that phonological knowledge becomes more fine-grained over time as phonological

representations become increasingly abstracted from the word forms of which they are a part.

Given the apparent contributions of both phonological working memory and long-term linguistic knowledge to NWR performance, it is perhaps surprising that in the NWR literature there is very little discussion of how these two processes interact. There is certainly a variety of research in NWR related domains that explain how short-term memory and long-term knowledge may combine. For example, Botvinick and Plaut (2006) show how performance on a variety of immediate serial recall tasks can be explained by how patterns of weights across units in a recurrent neural network (the model's long-term knowledge of the task) interact with task relevant aspects (the short-term activations across the units in the network). Immediate serial recall has strong links with NWR since both require one to immediately recall a sequence of stimuli. Gupta and Tisdale (2009) therefore adapted the Botvinick and Plaut model to apply it to NWR. The model maintains long-term knowledge of the syllables of words and nonwords together with the context in which they appeared. Again, short-term memory is represented as the set of activations across the units in the network. Although the model is not presented with naturalistic data as input (the training involves individual words and nonwords), the model does show how repetition performance can be influenced by interactions between short-term memory and long-term knowledge.

Verbal theories of NWR performance tend to lag behind these computational accounts. For example, although the phonological working memory account of NWR now includes redintegration (Schweickert, 1993) to cater for long-term influences, little thought has been given to its use in NWR. Redintegration uses phonological

knowledge to help 'fill in' any gaps when the representation of a nonword becomes

compromised in phonological working memory (e.g., Gathercole, 1995, 2006; Thorn,

Gathercole, & Frankish, 2005; see also Hulme et al., 1997). However, studies have

shown that wordlikeness effects emerge even for short nonwords that would not be

expected to tax memory capacity (e.g., Briscoe, Bishop, & Norbury, 2001; Jones,

Gobet, & Pine, 2007).

Explanations based on chunking

One theory aimed at specifying the link between short-term memory and long-

term memory is the chunking hypothesis (Miller, 1956). Chunking refers to the

continuous grouping and recoding (chunking) of a sequence or pattern of meaningful

stimuli based on one's exposure to those stimuli. Long-term memory therefore

contains chunks, and short-term memory is constrained by the number of chunks that

it can keep active. Arguably, the computational models discussed earlier are

instantiations of a chunking account because they share two key attributes with

chunking. First, long-term patterns of activation gradually change with exposure to a

given input in a similar way to how a set of items may gradually become grouped into

a single chunk. Second, a specific pattern of long-term activation within a network

represents a particular piece (or chunk) of knowledge.

Although Miller initially viewed chunking as a conscious learning mechanism, it

is now thought of as an automatic learning process that applies across a range of

stimuli (Gobet, Lane, Croker, Cheng, Jones, Oliver, & Pine, 2001; Servan-Schreiber

& Anderson, 1990; Simon, 1974). Besides a plethora of literature that shows how

learning may be governed by chunking (e.g., Bartram, 1978; Egan & Schwartz, 1979;

Gobet et al., 2001), chunking has also been captured computationally in models such

as CHREST (Gobet & Simon, 2000), EPAM-VOC (Jones, Gobet, & Pine, 2007),

MOSAIC (Freudenthal, Pine, & Gobet, 2006) and TRACX (French, Addyman, &

Mareschal, 2011).

The phoneme is generally considered to be the smallest unit of spoken language

(e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Gervain & Mehler, 2010), but the

vast majority of lexical items are sequences of phonemes (e.g., mʌmiː, "mummy";

kæt, "cat") that are usually expressed as part of longer utterances (e.g., /wɒts θæt/,

"what's that?"; /wɛːrz θə bɔːl/, "where's the ball?"). The chunking of sound

patterns in spoken language is therefore likely to result in progressively longer

sequences of phonemes that reflect the frequency with which these sequences are

encountered in the input[1]. For example, a high frequency sequence such as "mummy"

will initially be encoded using one chunk for each phoneme, but repeated exposure

will result in the learning of chunked phoneme sequences such as /mʌ/, /ʌm/, /miː/,

/mʌm/, /ʌmiː/, and the lexical item /mʌmiː/. When applying chunking to NWR,

nonwords, by definition, will not exist as chunked phoneme sequences. However,

knowledge of chunked sub-lexical sequences like /mʌ/ and /ʌmiː/ will play a major

role in repetition ability.

The chunking hypothesis also assumes a limitation in short-term memory

capacity that restricts the number of chunks that can be held in memory at any one

time. For Miller (1956) this was 7+/-2 items, though more recent research suggests

---

[1] In the discussion, we consider the opposing view that words are initially learnt holistically.

that this figure may be considerably lower (Cowan, 2000; Gobet & Clarkson, 2004).

A spoken word, phrase, or utterance will be encoded using one or more chunks, with

the number of chunks influencing the extent to which the word, phrase, or utterance

can be remembered. For example, the child who has learnt the chunked phoneme

sequences /wɛːrz/ and /mʌmiː/ can encode the utterance "where's mummy" using

only two chunks. The child who has only learnt the chunked phoneme sequences /w/,

/ɛː/, /rz/, /mʌ/, and /miː/ will require five chunks to encode the same phrase. The

former child is therefore more likely to accurately represent the utterance in short-

term memory than the latter child. Similarly, nonwords that are encoded using few

chunks are more likely to be repeated accurately than nonwords that are encoded

using many chunks.

Crucial to the chunking account is the linguistic input that the child receives,

because this determines the amount of linguistic knowledge the child learns. When

the linguistic input is extensive and varied, the child will learn a large number of

chunked lexical and sub-lexical phoneme sequences, which will subsequently aid

their language performance. When the linguistic input is more restricted and

homogeneous, relatively few chunked phoneme sequences will be learnt.

Testing the chunking hypothesis: Children's NWR performance

The advantage of the chunking explanation of NWR is that it not only specifies

how long-term phonological knowledge and phonological working memory interact,

but it is also able to make predictions about how NWR performance will change

based on the composition of the nonwords. For example, nonwords that contain

lexical items and morphological markers (that would be expected to exist as chunks) should be repeated more accurately than nonwords that do not contain lexical items and morphological markers. Short nonwords should be repeated more accurately than long nonwords because the latter will be represented in phonological working memory using a greater number of chunks.

We therefore test the chunking hypothesis on three sets of nonwords that vary in their degree of wordlikeness. The first set is the 2-4-syllable nonwords of the Children's Test of Nonword Repetition (CNRep, Gathercole, Willis, Baddeley, & Emslie, 1994). The nonwords are high in wordlikeness because most of them contain lexical and morphological elements (Archibald & Gathercole, 2006; Jones, Tamburelli, Watson, Gobet, & Pine, 2010; Thal, Miller, Carlson, & Vega, 2005). The second set is the 3-syllable nonwords of Dollaghan, Biber and Campbell (1995), half containing a lexical item and half not, reflecting medium wordlikeness. The final set is a set of newly devised 3-syllable low wordlikeness nonwords that do not contain any lexical items or morphological markers.

Although the three chosen nonword sets allow us to test any interaction between short-term memory and long-term memory, the nonword sets do not allow a properly balanced design. This is because we purposely use two existing nonword sets that are very established in the literature. For example, use of the CNRep nonwords is widespread (e.g., Archibald & Gathercole, 2006; Briscoe, Bishop & Norbury, 2001; Conti-Ramsden, 2003). If we are able to show that repetition of existing nonword sets is heavily based on an interaction between short-term memory and long-term memory, then we are able to question explanations of nonword repetition that are primarily based on one or other of short-term memory and long-term memory.

The following predictions can be made for the nonword sets:

1.   When considering only 3-syllable nonwords, the highest repetition

accuracy will be for the high wordlikeness nonwords and the lowest

repetition accuracy will be for the low wordlikeness nonwords.

2.   For the high wordlikeness nonwords, repetition accuracy will decrease

as nonword length increases.

3.   For the medium wordlikeness nonwords, repetition accuracy will be

greater for those nonwords that contain a lexical item than for those that do

not.

Note that none of the verbal theories of NWR outlined above make all of these

predictions[2]. Those theories that are based on long-term knowledge only make

predictions 1 and 3, while those theories that are based on short-term memory

capacity only make prediction 2 (redintegration could help in terms of predictions 1

and 3, but it is unclear whether nonwords of three syllables become sufficiently

degraded in quality to warrant redintegration).

Method

---

[2] It is possible that these predictions would be captured by the Gupta and Tisdale (2009) model, but one
would need to run simulations in order to verify this.

Participants

25 children (5;4-6;8, *M* = 6;1; 10 male, 15 female) were recruited from three

primary schools in and around the Nottingham area. All children were English

monolinguals and had no hearing difficulties, as reported by their schoolteacher. All

children performed within the normal range for both language (as assessed by the

British Picture Vocabulary Scale-2, Dunn, Dunn, Whetton, & Burley, 1997) and

performance IQ (as assessed by the Wechsler Pre-school and Primary Scale of

Intelligence-3, Wechsler, 2004), and showed no speech difficulties (as assessed by the

Diagnostic Evaluation of Articulation and Phonology, Dodd, Hua, Crosbie, Holm, &

Ozanne, 2002).

Materials

High wordlikeness nonwords (Gathercole et al., 1994). This test comprised

30 2-4-syllable nonwords (we omitted the 5-syllable items due to 5-6 year old

children having difficulty repeating nonwords of this length). At each nonword

length, there were 10 nonwords. The nonwords were split into two blocks of 15 so

that each repetition session kept the child's attention.

Medium wordlikeness nonwords (Dollaghan et al., 1995). This test

comprised 6 pairs of 3-syllable nonwords, each pair varying by only one phoneme to

form either a syllable that was a lexical item (e.g., *bath*esis) or a nonsense syllable

(e.g., *fath*esis). One nonword set therefore contained nonwords composed entirely of

nonsense syllables and the other nonword set contained nonwords that each contained

a lexical item.

Low wordlikeness nonwords. This test comprised sixteen 3-syllable nonwords that did not contain any morphemes or lexical items that were known to the children (based on the Children's Printed Word Database [CPWD] of word frequencies for 5-9 year old children, Masterson, Stuart, Dixon, & Lovejoy, 2010).

Design

Across nonword sets, wordlikeness (high, medium, or low) was the within subjects variable. For the high wordlikeness nonwords, nonword length was the within subjects variable. For the medium wordlikeness nonwords, the within subjects variable was whether or not the nonword contained a lexical item. In all cases, the dependent variable was the accuracy of the repetition.

Procedure

All children were assessed on an individual basis in a quiet room within the school and away from their classroom. Testing normally comprised three separate 15-minute sessions spread across several days or weeks (depending on the school and availability of the children). Each session involved administering one psychometric test and one or two nonword tests. Presentation of the nonword files was counter-balanced. Repetition responses were recorded onto a Sony ICD-MX20 digital voice dictaphone for later transcription.

Results

Nonword repetitions were transcribed into their phonemic form by the fourth author. A random sample of 15% of the children's recordings was transcribed by a trained linguist and a researcher experienced in transcribing nonwords, but who was not working on this project. Average inter-rater reliability was 87.3% (range: 85.0%-92.3%).

Figure 1 shows the repetition accuracy for each of the nonword sets. For the analyses, we take each of our three predictions in turn. Prediction 1 related to wordlikeness effects that were expected to be present across the 3-syllable stimuli of the three different nonword sets. This was confirmed using a one-way ANOVA, $F(2,48) = 8.15$, $p = .001$, $\eta_p^2 = .25$. However, post hoc Bonferroni tests showed that while high wordlikeness nonwords were repeated more accurately than low wordlikeness nonwords ($p = .002$), there was no statistically significant difference between high wordlikeness and medium wordlikeness nonwords ($p = .138$) nor between medium wordlikeness and low wordlikeness nonwords ($p = .197$).

INSERT FIGURE 1 ABOUT HERE

Prediction 2 related to length effects that should be seen in the high wordlikeness nonwords. The high wordlikeness data showed a strong effect of nonword length, $F(2,48) = 26.66$, $p < .001$, $\eta_p^2 = .53$. Post hoc Bonferroni comparisons showed that repetition accuracy for 2-syllable nonwords was significantly higher than both 3- and 4-syllable nonwords ($p = .01$ and $p < .001$ respectively) and 3-syllable nonwords were repeated more accurately than 4-syllable nonwords ($p < .001$).

Prediction 3 related to lexical item effects in the medium wordlikeness

nonwords. Half of these nonwords contain a lexical item and should therefore be

repeated more accurately than the other half of the nonwords that do not contain a

lexical item. This prediction was not supported statistically, $t(24) = .64$, $p = .527$,

Cohen's $d = .13$.

It is clear that the predictions of the chunking account are only partially

supported in 5-6 year old children. One explanation for the pattern of results observed

is the reliance that chunking places on long-term phonological knowledge, because

long-term knowledge determines the extent to which a given linguistic input is

represented accurately in short-term memory. This is the main weakness of chunking

as a verbal theory: unless one is able to reasonably estimate phonological knowledge,

any predictions based on chunking are likely to lack precision. In fact, *this is the case*

for any verbal theory that emphasizes the role of long-term knowledge in task

performance.

In order to conduct a more precise test of the chunking hypothesis, our chief

requirement is therefore the ability to approximate children's phonological

knowledge. A secondary requirement is a precise specification of how phonological

knowledge interacts with a given linguistic input to constrain the extent to which the

input can be represented accurately in short-term memory.

The linguistic input that the child receives is known to be an important

determinant of their vocabulary acquisition (e.g., Hoff & Naigles, 2002; Huttenlocher,

Haight, Bryk, Seltzer, & Lyons, 1991). We can therefore expect the language that the

child hears to have a major influence on the phonological knowledge that the child

subsequently acquires. In order to begin to estimate the child's phonological

knowledge, we require the following:

(1) A large input set that approximates the kind of input received by 5-6 year

old children;

(2) A computational implementation of chunking that can use linguistic input to

acquire phonological knowledge using mechanisms that are analogous to those in the

verbal theory;

(3) A plausible account of short-term memory capacity and how it interacts with

phonological knowledge that can constrain learning and task performance.

The next section describes a computational implementation that meets these

requirements.

CLASSIC: A computational implementation of the chunking hypothesis

CLASSIC (Chunking Lexical and Sub-lexical Sequences in Children) is a

computational model of the chunking process that is indistinguishable from the model

previously labelled EPAM-VOC (Jones, 2012; Jones, Gobet, & Pine, 2007, 2008).

This model was previously labelled EPAM-VOC because it is based on the

Elementary Perceiver and Memorizer (EPAM, Feigenbaum & Simon, 1984). It is

renamed here in order to make explicit the relation between the model and the account

that it is intended to implement. CLASSIC uses large corpora of naturalistic speech

data as input and learns long-term chunks of phonological knowledge[3]. Both chunk

---

[3] Other models of NWR exist such as OSCAR (Brown, Preece, & Hulme, 2000), the primacy model
(Page & Norris, 1998), and a model of phonological vocabulary learning (Gupta & Tisdale, 2009).

learning and NWR performance are constrained by a short-term memory capacity that

interacts with phonological knowledge. When a large amount of phonological

knowledge exists, all or a large proportion of a given input utterance or nonword is

likely to be processed; when only a small amount of phonological knowledge exists,

only a small proportion of a given input utterance or nonword is likely to be

processed. We will now describe the large-scale naturalistic input set that is used in

the model, together with the way that long-term chunks are learned, how they interact

with short-term memory capacity, and how the model performs nonword repetition.

Large-scale naturalistic datasets

    CLASSIC begins with no phonological knowledge except the individual

phonemes of the English language. From this starting state, the model must learn

phonological knowledge in a way that approximates children's learning. Variation in

the input that children receive at different ages therefore needs to be reflected in the

model's input. For example, the linguistic input a 5-6 year old child receives is

somewhat different from the input recived by a 2-3 year old child. We therefore

employ two input sets, one relating to 2-3 year old children ('younger input') and the

other relating to 4-6 year old children ('older input'). Over time, the model gradually

receives a higher proportion of the older input at the expense of the younger input.

    For the younger input, we use the maternal utterances from the Manchester

corpus (Theakston, Lieven, Pine, & Rowland, 2001) taken from the CHILDES

None of these models uses large corpora of naturalistic data as input, which we have argued is critical
in estimating long-term phonological knowledge.

Running head: VERBAL THEORIES AND COMPUTATIONAL

IMPLEMENTATIONS

Page 18

database (MacWhinney, 2000). The Manchester corpus consists of 12 sets of mother-child interactions involving 2-3 year old children, recorded over the course of a year and including an average of 25,519 maternal utterances (range 17,474–33,452).

For the older input, parental utterances to 4-5 year-old children are used (again taken from CHILDES) together with a large set of sentences from story books that are aimed at 5-6 year-old English children (e.g., "Snow White" and "The Ugly Duckling").

All of the input is converted into its phonemic equivalent using the CMU Lexicon Database (available at *http://www.speech.cs.cmu.edu/cgi-bin/cmudict*). The model is run independently for each of the 12 mothers, meaning that a fresh model is used with each set of mother's utterances. The advantage of this approach is that it results in variation in the model's NWR performance. The amount of input seen by the model is the number of utterances that are produced by each mother. For example, Anne's mother produces 31,393 utterances, and therefore the Anne model is presented with 31,393 utterances.

When the model begins learning, 100% of its input is based on a random selection of utterances (without replacement) from the younger input set. As learning proceeds, an increasing proportion of the younger input is replaced with the older input, again randomly selected without replacement. For example, at the beginning of the 31,393 utterances presented to the Anne model, 100% are taken from Anne's mother. After 10% of the input has been seen (3,139 utterances) the input changes so that 90% is from Anne's mother and 10% is from the 'older' input set. After the next 10% has been seen (6,279 utterances), 20% of the input is from the 'older' set and

80% from the 'younger' set. This pattern is repeated until 100% of the input has been presented.

Note that there is a random element to the model. It is therefore run 10 times for each mother. As we will see below, the repetition process for the model also has a random element to it, and hence every NWR test is administered to the model 10 times. The simulations therefore produce a total of $12 \times 10 \times 10 = 1,200$ NWR results for each of the NWR tests.

For statistical analyses, only 2 sets of NWR results are used from each mother (2×12 mothers = 24 sets of NWR results). The model runs from each mother have a great deal of overlap between them because they are derived from the same mother input; similarly the 'older' input across all of the models is sampled from the same source. The variance across all 1,200 model runs is therefore likely to be very low. Using only 2 runs per mother enables us to match the sample size of the children while keeping the overlap in input across the models to a minimum. Two model runs are therefore taken from each mother that are representative of the results for that mother (having a repetition accuracy within +/- 5% of the average repetition accuracy across all of the 100 runs for that mother).

Learning long-term phonological knowledge

Phonological knowledge is represented in CLASSIC by chunked phoneme sequences that are both sub-lexical (e.g., /mʌ/) and lexical (e.g., /mʌm/). Any given

input to the model is encoded using as few chunked phoneme sequences as possible[4].

For example, when the model has yet to begin learning, the representation of /wɒt/

("what") will comprise three chunks (one for each of the constituent phonemes).

However, later learning may include the chunk /wɒ/ and therefore the same input can

now be encoded using two chunks, /wɒ/ and /t/. Chunked phonological knowledge is

therefore critical for encoding an input using the smallest number of chunks, a process

that is important when we consider short-term memory capacity later.

The process by which chunks are learnt is very simple. Once an input is

represented in as few chunks as possible, CLASSIC learns a new chunk for each

adjacent set of chunks. If the input was /gəʊɪŋ/ ("going") and was encoded using the

two chunks /gəʊ/ and /ɪŋ/, a new chunk would be learnt for the whole word.

Similarly, if the input was /kæt/ and was encoded using only the individual

phonemes, the chunks /kæ/ and /æt/ would be learnt. Although chunk learning may

appear to occur rather too quickly, it is important to realize that this reflects the large

difference in scale between the amount of input received by the model (which

includes only a very small subset of the language to which the child is exposed) and

the amount of input received by the language-learning child.

The interaction between long-term memory and short-term memory

---

[4] When an input can be represented in different ways using the same number of chunks (e.g., /b eɪ k/ and /ɪ ŋ/ or /b eɪ/ and /k ɪ ŋ/ for 'baking'), the representation containing the highest frequency chunks is selected. Outside of the model, an articulation process uses frequency on the assumption that frequently encountered chunks will be easier to articulate than infrequent chunks.

A variety of research suggests not only that children's auditory processing is limited by short-term memory capacity (e.g., Baddeley, 1986; Gathercole & Baddeley, 1989), but also that this capacity is time-limited (e.g., Baddeley, 1986; Baddeley & Hitch, 1974). We therefore represent short-term memory capacity using a temporal duration of 2,000 ms (Baddeley, Thompson, & Buchanan, 1975) and allocate a time to match each of the chunks that are used to represent the input.

The time to match a chunk is based on estimates from Zhang and Simon (1985) who suggest that it takes 400 ms to match a chunk and an additional 30 ms to match each phoneme in a chunk except the first. In this way, (word) length effects are produced since a greater amount of time is required to match a chunk that contains many phonemes as opposed to a chunk that contains few phonemes. For example, the chunk /kæt/ would require 460 ms and /mʌmiː/ would require 490 ms. The time allocated to a given input is the sum of the times to match each of the chunks that represent the input.

When a given input has an allocated time that is below the 2,000 ms temporal duration of short-term memory capacity, it can be represented accurately within short-term memory. However, when a given input has an allocated time that exceeds 2,000 ms, the representation of the input in short-term memory is compromised. This means that fewer adjacent chunks can be learned. The probability of learning a chunk is 2,000 ms divided by the time required to represent the input. For example, if the input is represented by five chunks each with a time allocation of 460 ms (resulting in a total access time of $5 \times 460 \text{ ms} = 2,300 \text{ ms}$), then the probability of learning a new chunk for each adjacent pair of chunks would be 2,000 ms / 2,300 ms = .87.

Although the mechanism by which short-term memory capacity and chunked phoneme sequences interact is simple, it is also powerful. When the model has learnt few chunks, any given input is likely to require many chunks to represent it – thus compromising the representation of the information held in short-term memory and leading to a low level of learning. When the model has learnt many chunks, it is likely that any given input can be represented with few chunks and this in turn will result in a high level of learning.

CLASSIC is consistent with – but more precisely specified than – the verbal chunking theory on which it is based. However, to perform nonword repetition we need to consider an additional process: articulating the contents of a chunk. We explain this process in the next section.

Performing nonword repetition

NWR is achieved by presenting the model with the phonemic representation of each individual nonword in the same way that normal speech input is presented to the model. A nonword is therefore encoded using as few chunks as possible. Consistent with the interaction between short-term memory capacity and long-term chunked phoneme sequences, when the time allocation for the chunks exceeds 2,000 ms then the likelihood of accessing the contents of each chunk is probabilistic, using the same method as that outlined for learning chunks. However, an additional process is used in articulation because it is likely that the more often a sequence of phonemes has been used, the higher the probability of being able to articulate the sequence correctly. The frequency with which a chunk has been used during the course of the model's learning is therefore taken into account. This assumption is supported by good

correlations between the frequency with which phonemes and consonant clusters are

used by the children in the Manchester corpus and the age of acquisition of the

phonemes and clusters (from Smit, Hand, Freilinger, Bernthal, & Bird, 1990), $r(48) =$

-.51, $p < .001$. However, prior to assessing whether or not a sequence will be

articulated correctly, the frequency of a chunk is divided by the number of chunks that

are required to represent the input. An input that is represented using a small number

of relatively infrequent chunks (e.g., /ʌ g/ and /l iː/ for 'ugly') is likely to be easier

to articulate than the same input represented using a greater number of relatively

frequent chunks (e.g., /ʌ/, /g/, /l/, and /iː/) because the former indicates greater

experience with the phoneme sequences that are contained in the input[5].

The threshold for error-free articulation is a frequency of 10,000. For

frequencies below that, the probability of correct articulation is Log(chunk frequency)

/ Log(10,000). For example, when a chunk has only been accessed 100 times, the

probability of accurate repetition of its contents is .5; when the frequency is 1,000 the

probability is .75; and when the frequency is 5,000 the probability is .92.

Results

Figure 2 shows the repetition accuracy for the nonword sets in two formats: averaged

across the 2 runs per mother (2×12 mothers = 24 runs), and averaged across all 1,200

runs of the model. All statistical analyses are based on the 24 runs. The 1,200 run data

---

[5] In essence, this implements the assumption that exposure to a phoneme sequence is developmentally more advantageous than exposure to the component parts of the sequence in separate contexts, an assumption that dates back as far as Jakobson (1968/1941) but is also made more recently by Demuth, Culbertson, and Alter (2006) and Goad and Brannen (2003), amongst others.

is included simply to show that the 24 runs are representative of the overall

performance of the model.

Prior to testing the predictions outlined in the introduction, we show that the

model provides a good match to the child data. There is a strong correlation between

the child data and the 1,200-run model ($r(4) = .92$, $p = .010$) and between the child

data and the 24-run model ($r(4) = .91$, $p = .012$). We also calculate Root Mean

Squared Error (RMSE) on these figures. RMSE indicates the average discrepancy

between two datasets. For children and the 1,200-run model, the RMSE is 5.43,

indicating that on average, the 1,200-run model is within 5.43% of all of the

children's datapoints. The RMSE is 6.19 for children versus the 24-run model. In

combination, the correlation and RMSE results show that (a) the model provides a

very good fit to the children's data; and (b) the 24-run model is a good approximation

to the performance seen when all of the 1,200 simulations are considered.

Each of our three predictions is taken in turn, as per the children. A one-way

ANOVA confirmed the wordlikeness effects in the 3-syllable stimuli (prediction 1),

$F(2, 46) = 4.16$, $p = .022$, $\eta_p^2 = .15$), with post hoc Bonferroni tests indicating higher

levels of repetition accuracy for 3-syllable high wordlikeness nonwords over the low

wordlikeness nonwords ($p = .024$), with no further differences across groups ($p = .072$

or greater). These results mirror those of the children.

INSERT FIGURE 2 ABOUT HERE

The repetition accuracy data showed a strong nonword length effect for the high

wordlikeness nonwords (prediction 2), $F(2,46) = 50.64$, $p < .001$, $\eta_p^2 = .69$. Post hoc

Bonferroni comparisons showed that repetition accuracy for 2-syllable nonwords was significantly higher than for both 3- and 4-syllable nonwords, and 3-syllable nonwords were repeated more accurately than 4-syllable nonwords (all ps <= .001). These results also mirror those of the children.

In keeping with the child data, there was no difference between the medium wordlikeness nonwords that contained a lexical item and the ones that did not (prediction 3), $t(23) = 1.62$, $p = .118$, Cohen's $d = .33$.

These results show that, although the child data only partially support the predictions of the verbal chunking theory, they are nevertheless highly consistent with the predictions generated by CLASSIC, the computational implementation of the theory. The model is a faithful representation of the verbal theory. However, one key variable that is outside of the verbal theory is the nature of the linguistic input that the child receives, which is the source of the phonological knowledge that the child and the model learn. We can therefore examine the phonological knowledge in the model to see if it provides some clues as to the differences in the predictions made by the verbal theory and the computational implementation.

When the model is examined in more detail, it becomes apparent that two main factors influence repetition accuracy. First, encoding nonwords that are higher in wordlikeness requires fewer chunks than encoding nonwords that are lower in wordlikeness ($t(23) = 10.71$, $p < .001$, Cohen's $d = .92$). On average, 3-syllable high wordlikeness nonwords required 3.22 chunks (SD = .58) whereas low wordlikeness nonwords required 3.87 chunks (SD = .67). It is this difference that is the primary source of the wordlikeness effect.

Second, some of the lexical parts of nonwords do not have robust

representations. A good example of this is the medium wordlikeness nonwords,

because one set of these all have a lexical item in them. Table 1 shows the CPWD

frequency for each of the lexical items in the medium wordlikeness set of nonwords,

how often each item appears in a single chunk in the model across 1,200 simulations,

and the average frequency of the chunk in which each lexical item appears. It is clear

from the table that solid representations only occur for two of the lexical items: *bath*

and *kiss*. Not surprisingly, the nonwords containing these items have the highest

accuracy in the model (66% and 68% respectively).

INSERT TABLE 1 ABOUT HERE

Discussion

We have described chunking theory and how it applies to NWR performance,

the focus of the theory being the phonological knowledge that the child acquires. The

chunking account makes a series of predictions relating to length and wordlikeness

effects in the NWR performance of 5-6 year old children. A study using children of

this age showed that these predictions were only partially confirmed. Simulations

using CLASSIC, a computational implementation of chunking theory, showed that

when the linguistic input of 5-6 year old children was estimated in a reasonable way,

the model matched the child data in every respect. These results suggest that

phonological knowledge plays a critical role in NWR performance. The simulations

using CLASSIC show that the predictions of the chunking hypothesis can only be properly verified using a computational implementation. We argue that this is the case generally with verbal theories in psychology, but particularly for those that involve long-term knowledge.

One of the two primary aims of the current research was to highlight the role of chunking in NWR performance. This was accomplished by using three sets of nonwords that varied in their wordlikeness. In doing so, emphasis was placed on the interaction between short-term memory capacity and long-term phonological knowledge – thereby providing a stringent test of the chunking hypothesis. Once the linguistic input to 5-6 year old children had been estimated in a reasonable way, and the theory implemented as a computational model, a remarkable match to children's performance was shown across all three nonword sets. In terms of the goodness of fit between the model and child data, all correlations were extremely high and all RMSE rates were low. In terms of the accuracy of nonword repetition, the model showed exactly the same pattern of statistical effects as the children. A plausible account of the learning of phonological knowledge together with a reasonable account of the linguistic input the child receives provides a very good explanation of the child data.

The computational implementation of the chunking hypothesis clearly specifies how short-term memory capacity and long-term knowledge interact with one another. With learning, the effective size of short-term memory increases – as the *size* of the chunks increase, the same *number* of chunks can now capture a larger amount of information. An analysis of the number of chunks that were required to represent nonwords showed how differences in chunk size were the major reason for the

superior repetition accuracy of high wordlikeness nonwords over low wordlikeness

nonwords.

By exploring the intricacies of the model it was also possible to explain NWR

performance further. For example, both the children and the model failed to show the

same effects for medium wordlikeness nonwords that were seen in the 10-11 year old

children from Dollaghan et al. (1995). Dollaghan and colleagues found a significant

difference in repetition performance for nonwords containing a lexical item over ones

that did not, whereas no such difference was found in either our children or our

model. Detailed examination of the model suggested a straightforward explanation of

the discrepancy between the two studies: the linguistic input that 5-6 year old children

receive is not sufficiently rich to lead to wordlikeness effects for these nonwords at

this stage of language learning.

The second primary aim of the current study was to demonstrate the need to

develop computational implementations of verbal theories. The predictions that arose

from a verbal theory of chunking were not wholly borne out in the child data because

a large part of chunking theory concerns long-term knowledge. The CLASSIC model

of chunking was able to explain the child data in its entirety. The computational

implementation of chunking theory required not only a well-specified description of

how chunks were learnt and how they interacted with linguistic input and STM, but

also a well-specified description of the linguistic input itself.

Taken together, the predictions from the verbal theory of chunking and the

predictions from the computational implementation of chunking show how important

it is to be able to estimate a person's long-term knowledge. If one is unable to reliably

estimate this knowledge, any verbal theory involving long-term knowledge will fail to

make accurate theoretical predictions. In many cases, therefore, verbal theories will lack sufficient precision to make detailed predictions about experimental data. Implementing verbal theories as computational models such as CLASSIC – and thus clearly specifying concepts such as long-term knowledge – helps to produce more accurate theoretical predictions.

However, although CLASSIC provides a very good match to the child data, some researchers may take issue with some parts of the theory and its approach to language learning. In particular, the theory suggests that phonological knowledge is learnt by gradually building upwards from the phoneme. This contrasts sharply with the view that word learning is holistic from a relatively early age (Hallé & Boysson-Bardies, 1996). Supporting the holistic view, research on phonological awareness has shown that young children find it difficult to break words into constituent parts such as onset and rime (e.g. Carroll, Snowling, Stevenson, & Hulme, 2003; Liberman, Shankweiler, Fischer, & Carter, 1974). We do not think that these empirical data are inconsistent with chunking, for two reasons. First, awareness tasks explicitly test one's ability to manipulate word parts, whereas nonword tests are an implicit measure of phonological knowledge. Children may therefore be able to benefit from sub-lexical knowledge when repeating nonwords even though they cannot use this implicit knowledge in phonological awareness tasks. Second, the developing child rarely uses parts of words in everyday speech. Almost all of the spoken information for children therefore involves whole words and it may be the case that – even though knowledge of these words was gradually built up from individual phonemes and phoneme sequences – the words themselves have now become so ingrained as to be difficult to decompose back into sub-lexical sequences.

Like any scientific model, the model presented here abstracts away from and simplifies aspects of reality. Clearly, other processes beyond chunking are involved in NWR performance. For example, Coady and Aslin (2004) suggest that several additional processes are involved in NWR such as the perception and encoding of novel sound sequences (see also Jones & Witherstone, 2011), their temporary storage, and the articulatory mechanisms by which the sounds are reproduced. However, we believe that chunking is directly related to the most significant predictor of NWR performance – the child's existing phonological knowledge. While other factors may also produce shifts in performance in one direction or another, our view is that the child's phonological knowledge is the major influence on NWR performance.

In summary, we have presented a chunking theory of the learning of phonological knowledge together with a study of 5-6 year old children's NWR performance. Our results showed that, with reasonable linguistic input, a computational implementation of the chunking hypothesis was able to closely simulate the repetition data across three different nonword sets. In particular, the model was able to reproduce subtle effects of wordlikeness that depend on the detail of the stimuli used for training and testing the model. This illustrates both the importance of using input that is representative of the speech that children hear and the fact that computational models, but not verbal theories, can take advantage of such input to capture the statistical properties of the environment. CLASSIC incorporates a theory of how short-term memory capacity and long-term phonological knowledge interact in the learning of novel sounds that is consistent with the child data presented. Together with its previous incarnation EPAM-VOC, CLASSIC can replicate a wide range of data on NWR with great precision. To our knowledge, no other model enjoys

such a high level of success with this task. The model shows how computational

implementations of verbal theories are required in order to precisely specify different

aspects of the theory. Together, the theory and model represent a step forward in our

understanding of children's NWR performance and provide valuable insights into

how children learn new words.

References

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research, 49*, 970-983.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47-90). New York, NY: Academic Press.

Baddeley, A. D., Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 14*, 575-589.

Bartram, D. J. (1978). Post-iconic visual storage: Chunking in the reproduction of briefly displayed visual patterns. *Cognitive Psychology, 10*, 324-355.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 37*, 391-403.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*, 201-233.

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate

sensorineural hearing loss and those with Specific Language Impairment. *Journal of Child Psychology and Psychiatry, 42*, 329-340.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127-181.

Carroll, J. M., Snowling, M. J., Stevenson, J., & Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39*, 913-923.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*, 867-873.

Coady, J. A., & Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology, 89*, 183-213.

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with Specific Language Impairment (SLI). *Journal of Speech, Language, and Hearing Research, 46*, 1029-1037.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for Specific Language Impairment (SLI). *Journal of Child Psychology and Psychiatry, 42*, 741-748.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech, 49*, 37-173.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne (2002). *Diagnostic Evaluation of Articulation and Phonology*. London, UK: Harcourt Assessment.

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics, 16*, 211-222.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale 2*. Windsor, UK: NFER-Nelson.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*, 421-436.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition, 7*, 149-158.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in early infancy. *Science, 171*, 304-306.

Feigenbaum, E.A. & Simon, H.A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8*, 305-336.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review, 118*, 614-636.

Freudenthal, D., Pine, J., & Gobet, F. (2006). Modelling the development of children's use of optional infinitives in English and Dutch using MOSAIC. *Cognitive Science, 30*, 277-310.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513-543.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*, 83-94.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200-213.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2*, 103-127.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61,* 191-218.

Goad, H., & Brannen, K. (2003). Phonetic evidence for phonological structure in syllabification. In Joost van de Weijer, Vincent J. van Heuven & Harry van der Hulst (Eds.), *The phonological spectrum* (pp. 3-30). Amsterdam: John Benjamins.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four … or is it two? *Memory, 12*, 732-747.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I. & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5,* 236-243.

Gobet, F. & Simon, H.A. (2000). Five seconds or sixty: Presentation time in expert memory. *Cognitive Science, 24*, 651-682.

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language, 61*, 481-502.

Hallé, P. A., & Boysson-Bardies, B. de. (1996). The format of representation of

recognized words in infants' early receptive lexicon. *Infant Behavior and

Development, 19*, 463-481.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon.

Child Development, 73, 418-433.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early

vocabulary growth: Relation to language input and gender. *Developmental

Psychology, 27*, 236-248.

Jakobson, R. (1968/1941). *Child language, aphasia, and phonological

universals.* [*Kindersprache, Aphasie, und allgemeine Lautgesetze* (Uppsala:

Almquvist & Wiksell)]. Translated by A. R. Keiler (The Hague: Mouton).

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., & Pine, J. M. (2008). Computer simulations of

developmental change: The contributions of working memory capacity and long-term

knowledge. *Cognitive Science, 32,* 1148-1176.

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-

term memory: A computational model of the learning of new words. *Developmental

Science, 10,* 853-873.

Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010).

Wordlikeness and frequency in specific language impairment: Accuracy and error

data from two nonword repetition tests. *Journal of Speech, Language, and Hearing

Research, 53,* 1642-1655.

Jones, G., & Witherstone, H. L. (2011). Lexical and sub-lexical knowledge influences the encoding, storage, and articulation of nonwords. *Memory & Cognition*, *39*, 588-599.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk (3rd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Masoura, E. V. & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*, 422-429.

Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children's printed word database: Continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology, 101*, 221-242.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3-19.

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.

Munson, B., Edwards, J., & Beckman, M. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*, 61-78.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*, 1033-1047.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761-781.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1019-1034.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*, 168-175.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 592-608.

Service, E. (1992). Phonology, working memory and foreign-language learning. *Quarterly Journal of Experimental Psychology, 48*, 98-107.

Simon, H. A. (1974). How big is a chunk? *Science, 183*, 482-488.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*, 779-798.

Stokes, S. F., Wong, A. M-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment:

The case of Cantonese. *Journal of Speech, Language, and Hearing Research, 49*, 219-236.

Thal, D. J., Miller, S., Carlson, J., Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language, and Hearing Research, 48*, 1481-1495.

Theakston, A. L., Lieven, E. V. M., Pine, J. M. & Rowland, C. F. (2001). The role of performance limitations in the acquisition of Verb-Argument structure: An alternative account. *Journal of Child Language, 28*, 127-152.

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology, 50*, 133-158.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*, 47-62.

Wechsler, D. (2004). *The Wechsler Preschool and Primary Scale of Intelligence 3 (UK).* London, UK: Harcourt Assessment.

Weismer, S., Tomblin, J. B., Zhang, X. Y., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech Language and Hearing Research, 43*, 865-878.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and

idioms: Chunking and acoustical loop hypothesis. *Memory and Cognition, 13*, 193-

201.

Table 1. CPWD frequency for nonwords containing a lexical item in the medium

lexicality nonword set. Also shown is how often the lexical item appears within a

single chunk in the 1,200 simulations, and if so, the frequency in CLASSIC of the

chunk in which the lexical item appears.

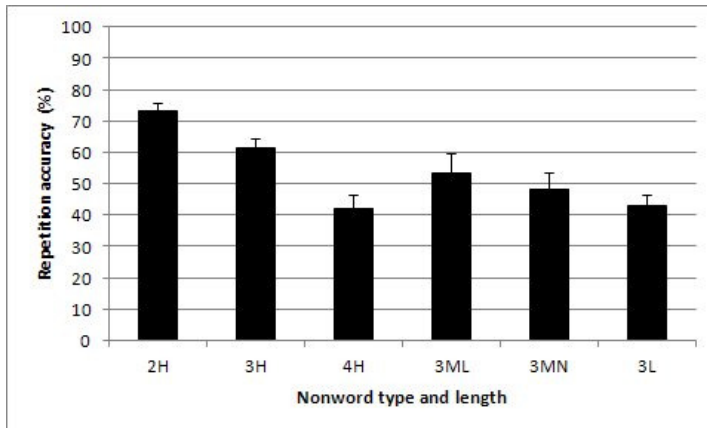| Lexical item	CPWD frequency	How often lexical item appears within a single chunk (%)	Average frequency of chunk in which item appears
Bath	257	82	108
Blame	22	23	201
Kiss	43	84	191
Ref	8	1	23
Speed	27	9	108
Trash	0	12	85

Figure 1. Nonword repetition accuracy (%) for the high wordlikeness nonwords, the

medium wordlikeness nonwords, and the low wordlikeness nonwords. Error bars

indicate standard error. Label key for x-axis: The numerals indicate the number of

syllables in the nonword; the letters indicate the nonword type (H = High

wordlikeness nonwords; ML = Medium wordlikeness nonwords containing a lexical

item; MN = Medium wordlikeness nonwords not containing a lexical item; L = Low

wordlikeness nonwords). For example, 2H represents 2-syllable high wordlikeness
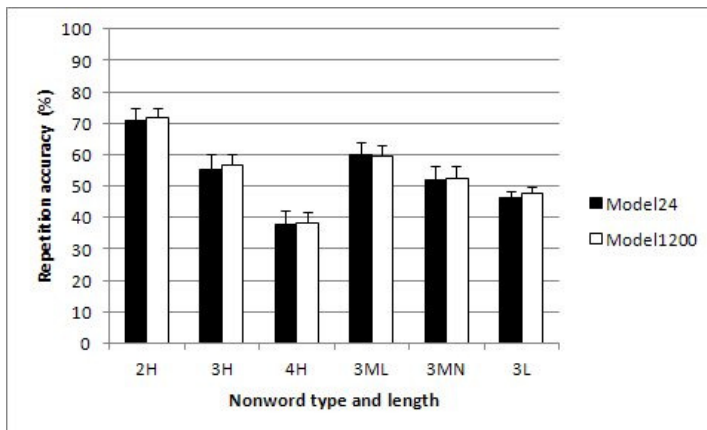
nonwords.

Figure 2. Nonword repetition accuracy (%) for the high wordlikeness nonwords, the medium wordlikeness nonwords, and the low wordlikeness nonwords, for the 24 representative runs of the model and all 1,200 runs of the model. Error bars indicate standard error. Label key for x-axis: The numerals indicate the number of syllables in the nonword; the letters indicate the nonword type (H = High wordlikeness nonwords; ML = Medium wordlikeness nonwords containing a lexical item; MN = Medium wordlikeness nonwords not containing a lexical item; L = Low wordlikeness nonwords). For example, 2H represents 2-syllable high wordlikeness nonwords.