Journal of Memory and Language 61 (2009) 481–502

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate

Prahlad Gupta *, Jamie Tisdale

Department of Psychology, University of Iowa, Iowa City, IA 52242, USA

ARTICLE INFO

Article history:
Received 7 November 2006
Revision received 28 July 2009
Available online 26 September 2009

Keywords:
Nonword repetition
Vocabulary learning
Computational modeling
Phonological memory
Word learning
Short-term memory
Language

ABSTRACT

The relationship between nonword repetition ability and vocabulary size and vocabulary learning has been a topic of intense research interest and investigation over the last two decades, following the demonstration that nonword repetition accuracy is predictive of vocabulary size (Gathercole & Baddeley, 1989). However, the nature of this relationship is not well understood. One prominent account posits that phonological short-term memory (PSTM) is a causal determinant both of nonword repetition ability and of phonological vocabulary learning, with the observed correlation between the two reflecting the effect of this underlying third variable (e.g., Baddeley, Gathercole, & Papagno, 1998). An alternative account proposes the opposite causality: that it is phonological vocabulary size that causally determines nonword repetition ability (e.g., Snowling, Chiat, & Hulme, 1991). We present a theory of phonological vocabulary learning, instantiated as a computational model. The model offers a precise account of the construct of PSTM, of performance in the nonword repetition task, of novel word form learning, and of the relationship between all of these. We show through simulation not only that PSTM causally affects both nonword repetition accuracy and phonological vocabulary size, but also that phonological vocabulary size causally affects nonword repetition ability. The plausibility of the model is supported by the fact that its nonword repetition accuracy displays effects of phonotactic probability and of nonword length, which have been taken as evidence for causal effects on nonword repetition accuracy of phonological vocabulary knowledge and PSTM, respectively. Thus the model makes explicit how the causal links posited by the two theoretical perspectives are *both* valid, in the process reconciling the two perspectives, and indicating that an opposition between them is unnecessary.

Introduction

Over the last two decades, the relationship between phonological short-term memory (PSTM) and vocabulary learning has become a major focus of investigation in psychological research, generating extensive bodies of study in the traditional domains both of memory research and of language research (e.g., Dollaghan, 1987; Gathercole & Baddeley, 1989; Gathercole, Service, Hitch, Adams, & Martin, 1999; Gathercole, Willis, Emslie, & Baddeley, 1992; Gupta, MacWhinney, Feldman, & Sacco, 2003; Martin & Saffran, 1997; Martin, Saffran, & Dell, 1996; Montgomery, 2002; Saffran, 1990; for review, see Baddeley, Gathercole, & Papagno, 1998; Gathercole, 2006). Among the results that initiated these bodies of research were the findings that novel word repetition ability (i.e., the ability to immediately repeat possible but nonoccurring word forms, also termed nonwords) is correlated with immediate serial list recall ability on the one hand, and with vocabulary achievement on the other, in normally developing children (Gathercole & Baddeley, 1989) and in children with specific language impairment (SLI; Gathercole & Baddeley, 1990a). Since these initial reports, an overwhelming amount of

* Corresponding author. Fax: +1 319 335 0191.
E-mail address: prahlad-gupta@uiowa.edu (P. Gupta).

0749-596X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2009.08.001

further evidence has documented the existence of a relationship between vocabulary size and/or new word learning; nonword repetition; and immediate serial recall (e.g., Atkins & Baddeley, 1998; Baddeley, 1993; Baddeley, Papagno, & Vallar, 1988; Gathercole & Baddeley, 1990b; Gathercole, Frankish, Pickering, & Peaker, 1999; Gathercole, Hitch, Service, & Martin, 1997; Gathercole et al., 1992; Gupta, 2003; Gupta et al., 2003; Michas & Henry, 1994; Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1992; Service, 1992; Service & Kohonen, 1995).

The question arises, of course, of what these associations indicate. One interpretation of these findings (e.g., Baddeley et al., 1998) has been that PSTM, the memory system hypothesized to underlie immediate serial recall performance, also underlies nonword repetition performance and novel word form learning. That is, in this view, PSTM plays a causal role in vocabulary learning, particularly phonological vocabulary learning. If this view is correct, the behaviorally observed patterns of relationship between immediate serial recall, nonword repetition, and vocabulary learning would be the manifestation of a mechanistic connection between the development of linguistic representations and what has been thought of as a memory system, and would have considerable importance for our understanding of language learning. It would also have considerable importance to the study of human cognition more generally, not merely because of the importance of language to cognition, but also because it would constitute a paradigm case of the interaction of short-term memory systems and long-term learning.

The perspective described above (e.g., Baddeley et al., 1998; Ellis & Beaton, 1993; Ellis & Sinclair, 1996; Gathercole, 2006; Gupta, Lipinski, Abbs, & Lin, 2005; Michas & Henry, 1994; Service & Craik, 1993) is based on the premises that (a) novel word or nonword repetition is a task that requires PSTM, much as immediate serial list recall does, and (b) learning a new word relies on processes that encompass those necessary for novel word or nonword repetition. The first premise is based on the extensive evidence indicating a relationship between nonword repetition ability and performance in immediate serial recall (the canonical PSTM task), including the finding of serial position effects in repetition of individual polysyllabic nonwords, analogous to those classically obtained in list recall (Gupta, 2005; Gupta et al., 2005). The task of nonword repetition is therefore, in this view, a means of gauging the efficiency of PSTM function. The starting point for the second premise is the observation that a novel word is, to a specific learner, in effect a nonword (Gathercole, 2006; Gupta, 2005; Gupta et al., 2005); it would therefore appear uncontroversial that the eventual learning of the novel word would depend on how well it could be *processed* as a novel word or nonword. From these two premises, it follows that PSTM is a causal determinant of vocabulary learning. This conclusion receives additional support from evidence that vocabulary learning and size are correlated with performance in immediate serial recall, the classic measure of PSTM (e.g., Gathercole et al., 1992; Gupta, 2003). This perspective is depicted graphically in Fig. 1(a), which shows that PSTM is posited to play a causal role in determining performance in immediate serial recall, performance in nonword repetition, and phonological vocabulary learning. In this account, the pairwise correlations between nonword repetition performance, immediate serial recall performance, and vocabulary learning arise because of the "third variable" they share in common, namely, PSTM. Correlations of nonword repetition and immediate serial recall with vocabulary size arise because vocabulary size is causally determined by vocabulary learning, as also shown in the figure.[1]

While emphasizing the causal role of PSTM in phonological vocabulary learning in this manner, some proponents of this view have also noted that the resultant phonological vocabulary knowledge (and, more generally, linguistic knowledge) is itself likely to play a role in determining nonword repetition ability (e.g., Gathercole, 1995, 2006; Gupta, 1995, 1996; Gupta & MacWhinney, 1997; Gupta et al., 2005). This conclusion is based in part on evidence that nonword repetition accuracy is affected by the wordlikeness of the nonwords, which suggests that lexical-phonological knowledge influences nonword repetition (e.g., Dollaghan, Biber, & Campbell, 1995; Gathercole, 1995; Gathercole, Willis, Emslie, & Baddeley, 1991).

The view described above has not gone unchallenged, however, and there is debate over whether the observed patterns of relationship do in fact imply a causal role for PSTM in phonological vocabulary learning. An alternative perspective is that the association between nonword repetition on the one hand and phonological vocabulary learning and size on the other may be mediated by phonological vocabulary size (and its concomitant phonological knowledge) itself, rather than by PSTM (e.g., Bowey, 1996; Edwards, Beckman, & Munson, 2004; Ellis Weismer & Edwards, 2006; Metsala, 1999; Snowling, 2006; Snowling, Chiat, & Hulme, 1991). Snowling et al. (1991, p.372) argued, for example, that "... we can turn the argument about causation advanced by Gathercole and her colleagues on its head: children with good vocabulary knowledge are better able to cope with the processing demands of nonword repetition tasks than are children with poor vocabulary knowledge". As another example, Edwards et al. (2004, p. 434) have also emphasized the causal role of phonological vocabulary knowledge, suggesting that "[their] results support an account of acquisition in which the typically developing child gradually

[1] The causal link from phonological vocabulary *learning* to phonological vocabulary *size* is not specifically a postulate of this particular account but rather is a logical necessity that any account presumably must incorporate. The reason for inclusion of this link in the depiction is to clarify the relationship between terms that are sometimes used interchangeably. In particular, the terms "vocabulary learning" and "vocabulary size" are sometimes used as if indicating the same construct. In fact, however, these terms refer to different constructs that should be clearly distinguished. A measure of *vocabulary learning* gauges the operation of processes that lead to increases in *vocabulary size*. The terms "vocabulary", "vocabulary size" and "vocabulary knowledge" are also often used interchangeably, but these, too, refer to different constructs: *vocabulary* refers to the set of known words; *vocabulary size* refers to the cardinality of this set (the number of items in the vocabulary), and *vocabulary knowledge* refers to the information that is known about the members of this set. Having noted these various distinctions, we will nevertheless adopt the practice of ignoring the latter set, for ease of exposition. However, as acknowledgment that different constructs are being conflated, we will frequently use the terms *vocabulary/knowledge* or *vocabulary size/knowledge* to refer to vocabulary/vocabulary size/vocabulary knowledge.

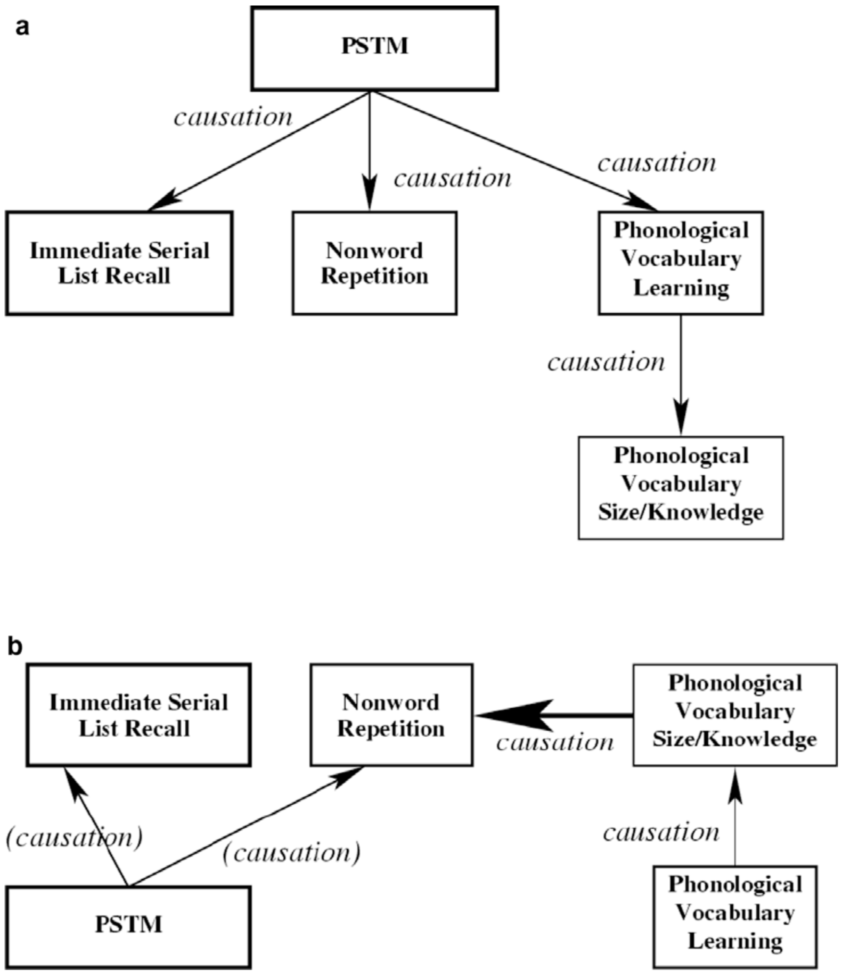

Fig. 1. Depiction of the two accounts. (a) The PSTM account. (b) The linguistic account.

acquires more and more robust phonological knowledge as a consequence of learning to produce many words. That is, an increase in vocabulary size does not simply mean that the child knows more words, but also that the child is able to make more and more robust phonological generalizations". (Such generalization would, in this view, underlie nonword repetition.)

This view is based on the premises that (a) factors other than PSTM influence performance in the nonword repetition task, and (b) one particularly relevant such factor is the nature and extent of phonological knowledge available to the individual. The first premise, which appears uncontroversial, is based on the intuition that performance of the nonword repetition task depends on multiple abilities, which may include PSTM, but are certainly not limited to PSTM. The second premise is based on the evidence cited above for the influence of lexical/phonological knowledge on nonword repetition accuracy (e.g., Dollaghan et al., 1995; Gathercole, 1995; Gathercole et al., 1991), as well as on evidence suggesting that as phonological knowledge changes with vocabulary development, it impacts nonword repetition ability. For instance, Munson, Kurtz, and Windsor (2005) reported that the effect of nonwords' phonotactic probability on childrens' accuracy in repeating those nonwords decreased as a function of vocabulary size, and interpreted this as further evidence that the level of phonological vocabulary knowledge causally affects nonword repetition performance. From these two premises, it follows that phonological vocabulary knowledge is a causal determinant of nonword repetition ability. Thus in this view, it is the extent of phonological vocabulary/knowledge available to an individual, and not PSTM, that plays the important causal role in the association between nonword repetition and vocabulary size.

This perspective is depicted graphically in Fig. 1(b), which shows that phonological vocabulary size/knowledge is posited to causally affect nonword repetition ability. In this view, the correlation between nonword repetition and vocabulary size is a direct consequence of this causal link, while the correlation between nonword repetition and vocabulary *learning* is mediated by the two causal links connecting them. In contrast with the preceding account, a causal link between PSTM and vocabulary learning is not emphasized here. This perspective does not usually include explicit hypotheses about immediate serial recall or PSTM, or their relationship to nonword repetition, but the causal links shown from PSTM to immediate serial recall and nonword repetition would not be inconsistent with the account. An explanation of the correlation between nonword repetition and immediate serial recall in terms of their shared PSTM third variable would therefore also be consistent with the account. An explanation of the correlation between vocabulary/knowledge and immediate serial recall in terms of the covariance arising from (i) the causal link from vocabulary/knowledge to nonword repetition and (ii) the shared role of PSTM in nonword repetition and immediate serial recall would not be inconsistent with the account.

In principle, these accounts are not mutually exclusive, as a number of investigators have noted (Bowey, 1996; Brown & Hulme, 1996; Gathercole, 1995, 2006; Gathercole, Service, et al., 1999; Gupta, 1995, 2006; Gupta & MacWhinney, 1997; Snowling, 2006; Snowling et al., 1991). For instance, the term *redintegration* (Schweikert, 1993) has been used to refer to a process whereby short-term memory recall is influenced by long-term knowledge representations, which aid in reconstruction of information from a short-term memory trace, and the idea of redintegration has been extended and applied specifically to nonword repetition by Gathercole et al. (1999), constituting an account that posits a role for both PSTM and long-term vocabulary knowledge in nonword repetition. As another example, Gupta's (1995, 1996; Gupta & MacWhinney, 1997) work has also argued that long-term linguistic knowledge inherently influences recall from short-term memory. Nevertheless, in practice, these two perspectives – one emphasizing the role of PSTM and the other emphasizing the role of long-term knowledge – have come to be somewhat oppositional; this continues to be the case, as evidenced by the extensive debate about these views in a recent review of the field and in commentaries accompanying that review (see Gathercole, 2006 and associated commentaries). In our view, an important reason for this has been the lack of an explicit, implemented computational account that can simulate the relevant phenomena and provide mechanistic understanding of whether/how the accounts could be consistent with each other.

We will refer to these opposing accounts, respectively, as the "PSTM account" and the "linguistic account" (corresponding roughly to what others have termed the "phonological storage hypothesis" and "phonological sensitivity hypothesis", respectively – e.g., Bowey, 1996; Gathercole, 2006). Let us consider carefully the points of agreement and disagreement between them. First of all, it may be noted that it is *phonological vocabulary knowledge* (rather than, say, semantic knowledge) that is particularly relevant to the hypotheses in both accounts. Second, both accounts posit or acknowledge the possibility of a role for the level of phonological knowledge in determining nonword repetition accuracy. The crucial point of difference is that the PSTM account also posits an important causal role for PSTM in the *growth* of phonological vocabulary/knowledge (i.e., in phonological vocabulary *learning*), while the linguistic account posits that such a role may be relatively minor. The critical discriminating question is, therefore, whether or not PSTM plays an important causal role in phonological vocabulary learning.

Although there have been numerous computational accounts of immediate serial list recall (e.g., Botvinick & Plaut, 2006; Brown et al., 2000; Burgess & Hitch, 1992, 1999, 2006; Gupta, 1996, 2009; Page & Norris, 1998) as well as of word and nonword repetition (e.g., Gupta, 1996, 2009; Gupta & MacWhinney, 1997; Hartley & Houghton, 1996; Vousden, Brown, & Harley, 2000), they have either not addressed phonological vocabulary learning, or have not addressed it in a manner that resolves the critical question (but see Jones, Gobet, & Pine, 2008 for a recent approach). The issues can, however, usefully be framed in computational terms, and in particular, in terms of three questions: First, what is required computationally for repetition of a novel word form? That is, what kind of computational mechanism would be necessary for such a task? Second, is PSTM part of these computational requirements? Third, do these necessary computational mechanisms and/or PSTM play any role in phonological vocabulary learning?

A useful starting point in thinking about these questions is the fact that a novel word form, on first exposure, is a novel sequence of sounds. The task of repeating such a stimulus immediately after exposure to it requires the listener to encode the serial order of this sequence during its presentation, and then replicate this serial order when the stimulus is no longer present. That is, immediate repetition of a novel word form requires the encoding and retrieval of a novel serial ordering of constituent sounds. What are the mechanistic underpinnings of a task that requires production of an unfamiliar serially ordered sequence?

Below, we present a theory of phonological vocabulary learning. We instantiate our theory as a recurrent neural network computational model, offering a precise account of the construct of PSTM, of performance in the nonword repetition task, of novel word form learning (i.e., phonological vocabulary learning), and of the relationship between all of these. We show through simulation not only that PSTM causally affects both nonword repetition accuracy and phonological vocabulary learning, but also that phonological vocabulary/knowledge causally affects nonword repetition ability. We further demonstrate that the model's nonword repetition performance shows effects of phonotactic probability and of nonword length, which have been taken as evidence for the causal effects on nonword repetition accuracy of phonological knowledge and PSTM, respectively. As a by-product of these demonstrations, we provide a means of reconciling the PSTM and linguistic accounts, suggesting that an opposition between them is misplaced.

Serial ordering within phonological word forms: A computational model

The model we constructed to examine these questions is adapted from recent work by Botvinick and Plaut (2006), who developed a neural network model of immediate serial recall of arbitrary lists of letters. The issue we investigated was not immediate serial list recall, but rather the processing, production, and learning of phonological sequences, where each such sequence constitutes a word

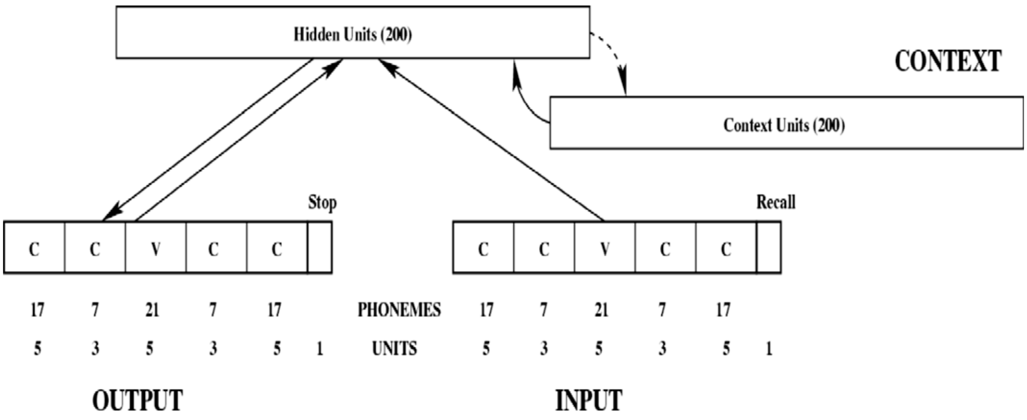

Fig. 2. Architecture of the model.

form. That is, the present model addresses serial ordering *within* word forms, investigating phonological sequencing and phonological vocabulary learning. It does not address serial ordering *across* word forms, as in list recall.

Following Botvinick and Plaut (2006), our model took the general form of a simple recurrent network (SRN), which is a type of neural network architecture that has been widely employed in simulation of cognitive phenomena involving sequential processing (e.g., Christiansen et al., 1998; Dell et al., 1993; Elman, 1990; Ernansky, Makula, & Benuskova, 2007; Gupta, 2009; Gupta & Cohen, 2002; Jordan, 1986). The architecture of the model is shown in Fig. 2. The model has an input layer at which a distributed representation of an entire syllable is presented, and an output layer that uses the same representation scheme, at which the model's output is produced. The representation of a syllable, at both the input and the output layers, is in terms of a CCVCC (i.e., Consonant–Consonant–Vowel–Consonant–Consonant) frame. That is, the representation scheme for a syllable consists of units divided into five slots. Activation of units in the first slot denotes the first C (if any) of the syllable, activation of units in the second slot denotes the second C (if any) of the syllable, activation of units in the third slot denotes the V of the syllable, and so on. Within each of these slots, the various phonemes that are legal for that slot for English are represented as different patterns of activations across a set of units. For example, for the encoding scheme used, there are 17 different phonemes of English that are legal for the first C slot. Five bits (i.e., binary digits) are needed to represent 17 different binary patterns, and five binary units were therefore necessary for the first slot. Thus each of the 17 possible phonemes was represented as a different binary pattern of activation across this set of five units. Similarly, the 21 phonemes that are possible in the V slot were represented as patterns of activation across a set of five units constituting the V slot (five units suffice for up to 32 different patterns); and so on for the various slots shown at the input and output layers in Fig. 2.

In addition to the input and output layers, the model has a hidden layer of 200 units. All units in the input layer project to all units in the hidden layer, and all units in the hidden layer project to all units in the output layer. The model also has a layer termed the context layer, as is typical of simple recurrent networks. Each context layer unit corresponds to one hidden layer unit; the context layer thus has the same number of units as the hidden layer. The dashed arrow indicates one-to-one connections from hidden layer units to context layer units. These connections are used to copy the hidden layer activation pattern into the context layer, at the end of each time step of processing. The context layer however has full connectivity *to* the hidden layer, and transmits its activation to the hidden layer in the same manner as does the input layer. At each point in processing, the model therefore has two types of input available: the current stimulus, available in the input layer; and, in the context layer, a copy of the hidden layer's previous activation pattern, which provides the network with a kind of memory of its previous processing. (It is worth noting that the *HiddenLayer → copy to ContextLayer → transmit to HiddenLayer* circuit is functionally equivalent to each hidden layer unit having connections back to itself and to all other hidden layer units. Therefore, for simplicity we will frequently refer to the *HiddenLayer–ContextLayer–HiddenLayer* circuit as the hidden layer self-connections or hidden layer *recurrent* connections. Also, we will use the term *context information* to refer to the information about previous processing that is provided by this circuit.) Finally, the model also incorporates feedback connections from the output layer back to the hidden layer, as in the Botvinick and Plaut (2006) model. The strength of all connections in the model can be adjusted via learning, with the exception of the one-to-one connections from hidden layer units to context layer units, which have fixed strength and implement the copy operation.

The model takes as its input a sequence of one or more syllables constituting a monosyllabic or polysyllabic phonological word form. That is, a word form is presented to the model one syllable at a time. For our primary goal of examining the role of PSTM in serially ordered production, it sufficed for the model to represent and produce serial order across syllables within a word form, but not within a syllable, as the model still would be processing and producing phonological sequences, and this was computationally simpler than attempting to represent word forms as, say, phoneme sequences. There is also considerable evidence to suggest that the syllable is a natural unit of phonological analysis, and that there is perceptual segmentation at the level of the syllable (e.g., Jusczyk, 1986, 1993, 1997; Massaro, 1989; Menyuk, Menn, & Silber, 1986) and it thus seemed reasonable to treat word forms as sequences of syllables rather than as sequences of phonemes. Each phoneme in each syllable was nevertheless represented individually, using the CCVCC scheme described previously. After the input had been presented, syllable by syllable, the model attempted to produce as its output the entire word form, as the correct sequence of syllables (including correct phoneme representations). Importantly, the model's output production was performed when there was no longer any information in the input about the word form that had been presented. In order to perform the task, the model therefore had to develop an internal representation of the entire word form that included information about both its serial ordering and the phonemic structure of the syllables comprising the word form. Over the course of many such learning instances, the model not only learned a phonological vocabulary (the set of phonological sequences that it could correctly produce) but also acquired generalized phonological knowledge (as we will demonstrate below).

As an example, let us consider the syllable-by-syllable processing of the word form *flugwish* in the model. The procedure is the same irrespective of whether the word form has been presented to the model previously. At the first time step, the first syllable *flug* is presented at the input layer. (At the start of processing a word form, the context layer is reset to zero; it therefore does not provide any input on this first time step.) In response to presentation of *flug* at the input, the model's task is to produce that same syllable at the output. The model's actual output, of course, may or may not be correct. Either way, after the model has produced an output, the actual output is compared with the target output (in this case, *flug*) and the discrepancy between the two is calculated (what in neural networks is often termed the *error*). Then, the activation pattern that is present at the hidden layer is copied to the context layer.

At the second time step of processing, the second syllable *wish* is presented at the input layer. The overall input to the network, however, is this representation of *wish* together with the context layer copy of the previous hidden layer activation pattern. The model's task, as for the first syllable, is to produce the current input syllable (now *wish*), at the output. Again, the output may or may not be correct, and the error is calculated. Again, the hidden layer activation pattern is copied to the context layer so as to be available at the next time step.

At the next (third) time step, however, there is no phonological input, because *wish* was the final syllable of the word form, so that presentation of the word form has been completed. At this point the model's task is to produce at the output layer the *entire* sequence of syllable representations previously presented at the input layer, i.e., *flug* followed by *wish*. That is, the network's task at this point is to repeat the preceding word form, as a serially ordered sequence. To indicate this to the network, the input is now a signal to recall the preceding sequence of inputs. This signal or cue is denoted by activation of the "Recall" unit in the input layer, as in Botvinick and Plaut (2006). It can be thought of as loosely analogous to the context signal that is present throughout recall in other types of models (e.g., Burgess & Hitch, 1999; Hartley & Houghton, 1996), although unlike those other models, it does not change during recall. Thus on the third time step, in the presence of the "Recall" input signal (together with the context layer input), the network's task is to produce *flug* at the output. This target and the network's actual produced output are again used in the calculation of error. Again, the hidden layer activation pattern is copied to the context layer so as to be available at the next (fourth) time step. At the fourth time step, the "Recall" signal is once again presented as input (together with context layer input). The network's task is to produce the second syllable of the preceding sequence, *wish*. Again, the difference between this target output and the actual output produced by the network is used to determine error. Again, the hidden layer activation pattern is copied to the context layer for availability at the next time step. At the next (fifth) time step, the input is still the "Recall" signal together with context layer input. The network's task on this time step (the final step in repetition of this word form) is to activate a specially designated "Stop" unit at the output layer, to signify the end of its production of the word form. Once again, this target and the network's actual produced output are used in the calculation of error. At the end of this repetition process, the total error that has been calculated over all five time steps is used to adjust the model's connection weights, using a variant of the back-propagation learning algorithm (*back-propagation through time*; Botvinick & Plaut, 2006; Rumelhart, Hinton, & Williams, 1986). For present purposes, the key aspect of this algorithm is that, by using error from all steps of processing, it adjusts weights on the connections from the context layer to the hidden layer in a way that takes into account information about the entire sequence whose presentation and attempted repetition has just ended.

Via connection weight adjustments following repeated exposures to a set of word forms in this manner, a neural network model of this kind would be expected to develop more and more detailed internal phonological representations of these word forms such that it could produce them with increasing accuracy. At each point in this trajectory of learning, the set of word forms that could be correctly produced would constitute the model's *phonological vocabulary*. The construction of a model that could learn a phonological vocabulary in this manner formed the starting point for the present investigations.

Simulations of phonological vocabulary learning and nonword repetition

Acquisition: Simulation 1

A set of approximately 125,000 phonologically distinct words of one through four syllables was drawn from the corpus of words accompanying the Festival speech synthesis software (Black & Taylor, 1997). A syllabified phonemic transcription was created for each word using a transcription scheme developed by Dennis Klatt (for description, see Luce & Pisoni, 1998). The transcription for each syllable in a particular word was then further translated into a set of binary vectors, one for each phoneme, according to the scheme described in the previous section.

The simulation of phonological vocabulary learning consisted in presenting the model with a set of 1000 words drawn from the overall set of 125,000, with adjustment of connection weights occurring after presentation of each word. This procedure (syllable-by-syllable presentation, and connection weight adjustment, for each of the 1000 words) was termed an *epoch*, and vocabulary learning consisted of a large number of such epochs (as discussed further shortly). The 125,000 words were intended to approximate the set of words of the language, and the sample of 1000 words in an epoch was intended to be very loosely analogous to the kind of exposure to a subset of these words of the language that a human learner might receive in a period of time such as a day. We are not aware of any estimates of the number of words that a human learner actually hears during a day, but it has been estimated that children speak 10–14,000 word tokens per day (Wagner, 1985), so that it does not appear unrealistic to assume exposure to 1000 words in a day. An epoch, as noted, was meant to correspond very roughly with such a period of time.

The 1000 words in each epoch were selected stochastically from the overall set of 125,000, with the probability of selection of a given word into the sample of 1000 for an epoch being based on its frequency of occurrence. Specifically, the probability of inclusion of a word was determined, as in Seidenberg and McClelland (1989), by the formula $P = K \times \log_{10}(kf + 2)$, where kf was the Kucera–Francis frequency of occurrence (Kucera & Francis, 1967), and K was a constant. Selection was with replacement within each epoch so that a particular word could occur more than once within an epoch.

For each of the words in an epoch, the procedure described in the previous section was followed: the model was exposed to the word one syllable at a time, attempting to repeat each syllable after its presentation, and attempting to repeat the entire word one syllable at a time after presentation was complete; connection weights were adjusted at the end of this process. Following presentation and connection weight adjustment for all 1000 words in an epoch, the model's production accuracy was then assessed in a test using the same procedure (but now without any adjustment of connection weights) one word at a time, for all 125,000 vocabulary items. For each of the five phoneme slots in the output layer, the model was considered to have produced the phoneme whose representation vector was closest to the actual activation pattern present in that slot. The model's performance in producing a given word was considered correct only if, when producing the entire word after its presentation, the model produced the correct phoneme in every slot of each syllable, and all the syllables in the correct serial order. Otherwise, the model's performance in producing that word was considered incorrect, and the word was not considered a correctly produced vocabulary item. The number of words correct in the test was taken as a measure of the model's phonological vocabulary size at that epoch. It may be worth clarifying that, in keeping with the aim of examining the learning and repetition of *phonological* word forms, the model incorporated no representations of meaning, i.e., no semantics. Thus, the model's phonological vocabulary consisted of the set of word forms to which the model had been exposed and that it could produce correctly following immediate presentation – i.e., phonological forms that it could repeat, rather than phonological forms that could be produced via semantics.

A further stochastic sample of 1000 words was then selected, and the entire above procedure repeated, constituting another epoch. Overall, 8000 epochs of such exposure-and-test were provided to the model, for a total of 8,000,000 learning trials. Keeping in mind the loose analogy of an epoch with a day, the 8000 epochs corresponds roughly with 8000/365 = 21 "years" of age, and keeping in mind the assumption of exposure to 1000 words per day in human language users, the figure of 8,000,000 exposures to words in 21 years would appear to be within the bounds of plausibility.

Fig. 3 shows development of the model's phonological vocabulary across these epochs. The upper curve in the figure shows the number of words to which the model has been exposed at least once. The lower line shows the size of the model's vocabulary, determined as above. As can be seen, the model's phonological vocabulary exhibits steady growth as a function of exposure to the simulated linguistic environment. By about 6000 epochs, phonological vocabulary size had largely asymptoted, with that level being maintained through the remainder of the 8000 epochs of training. In the absence of relevant empirical data on phonological word form repetition ability, we do not know how closely stages in the model's trajectory match developmental time points in the human trajectory. Clearly, however, the model exhibits a developmental trajectory; moreover, the trajectory exhibits the kind of power law learning often characteristic of human cognition.

Nonword repetition: Simulation 2

To test the network's ability to generalize its knowledge of trained word forms to the repetition of novel word forms, it was tested on a set of nonwords of lengths two through four syllables (100 nonwords of each length) by random sampling from a corpus of nonwords used in our behavioral studies (Gupta, 2005; Gupta et al., 2005; Gupta

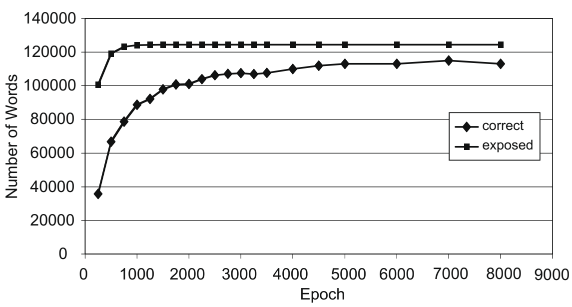

Fig. 3. Simulated phonological vocabulary learning.

et al., 2004). As described in greater detail in Gupta et al. (2004), all nonwords in this corpus have CV nonfinal syllables and a CVC final syllable, were constructed to be unwordlike, and contain no inflectional morphemes. To test nonword repetition, we presented each of the selected stimuli to the model, one at a time. Each stimulus was presented syllable by syllable, exactly as for vocabulary items in Simulation 1. As for vocabulary items, the model's repetition of the word form was tested immediately after its presentation. The model's repetition response was recorded, and was scored as correct only if all phonemes of each syllable were produced correctly, and all syllables were produced in correct serial order. Connection weights were not adjusted, to ensure that the model's performance in this task would be based solely on generalization from its existing phonological vocabulary/knowledge, rather than on any learning of the specific nonwords presented. Overall, this procedure was exactly like the test of production accuracy at the end of each epoch during learning of the vocabulary in Simulation 1, except that the stimuli being tested were not vocabulary items but nonwords.

As noted in Simulation 1, the model's phonological vocabulary size asymptoted by epoch 6000, and we therefore considered epochs 6000–8000 to constitute an ''adult'' range of phonological vocabulary achievement. We set the model's learning rate parameter so as to yield adult-like nonword repetition accuracy at this phonological vocabulary learning asymptote. The learning rate that achieved this was 0.0001, the same as that employed by Botvinick and Plaut (2006). Fig. 4 compares the model's nonword repetition performance at 8000 epochs, the asymptote of vocabulary acquisition, with that of human adult participants on nonwords of the same lengths drawn from the same corpus (Gupta et al., 2005), and confirms that performance at this final epoch in the model's adult range was quite similar to the human data. We also examined the model's nonword repetition performance at a number of points in this ''adult'' range, with similar results in all cases.

To further assess the model's nonword repetition performance, we analyzed error patterns in the model's nonword repetition data, comparing them with those in the behavioral data (Table 1). None of the model's parameters had been adjusted in any way to account for error patterns or any of the other phenomena described below. The model showed good correspondence with the behavioral patterns in a number of important respects. First, the overall proportion of syllables correctly repeated was very similar in the behavioral and the simulation data (94.97% vs. 95.56%). Second, the behavioral data show very strong adherence to the syllable-structure constraint, such that target syllables with CV structure were almost without exception repeated as CVs, even if the actual repetition was incorrect, and CVC syllables were repeated (even if incorrectly) as CVCs. The model shows equally strong adherence to this constraint, for both CVs and CVCs. (CVs and CVCs were the only two syllable types that occurred in the nonwords, in both the behavioral and the simulation experiments.) Third, we examined the proportion of incorrectly produced syllables that had a single-phoneme error versus multiple-phoneme errors. In the simulation results, as in the behavioral data, close to 90% of errors were single-phoneme errors. Finally, for single-phoneme errors, we further classified the error as a substitution (replacement of a phoneme in the target syllable by one that was not present in the target syllable), insertion (involving addition of a phoneme), or deletion (removal of a phoneme). In the model, as in the behavioral data, the overwhelming majority of single-phoneme errors were substitutions. The model thus exhibits a good correspondence with the behavioral data in this analysis of proportions of error types.

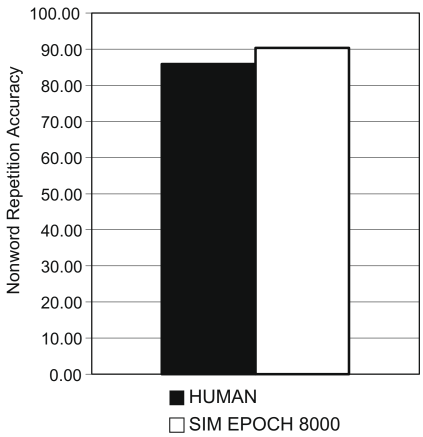

Fig. 4. Nonword repetition accuracy, collapsed across two-, three- and four-syllable nonwords. Comparison of behavioral data from Gupta et al. (2005) with simulation results at 8000 epochs.

Table 1
Error proportions in nonword repetition: comparison of behavioral and simulation data.

	Human (%)	Model (%)
Syllables correctly produced	95.0	95.6
Preserved syllable structure		
CV targets	99.7	98.9
CVC targets	99.4	98.2
Proportion of errors that are		
Single-phoneme errors	86.2	86.9
Multiple-phoneme errors	13.8	13.1
Single-phoneme errors		
Substitutions	90.1	79.5
Insertions	6.5	10.3
Deletions	3.4	10.3

Effects of nonword length and serial position

A robust phenomenon of nonword repetition has been the finding that repetition accuracy decreases with nonword length (e.g., Gathercole et al., 1992). This finding has been thought to indicate the operation of PSTM in nonword repetition both because it mirrors the well-documented finding of list length effects in immediate serial recall of lists, and because it appears difficult to account for in a purely linguistic account (e.g., Gathercole, 2006). A more recent finding is that serial position effects, which are likewise considered a hallmark of PSTM engagement in immediate serial list recall, are also obtained syllable-wise in repetition of single polysyllabic nonwords (Gupta, 2005; Gupta et al., 2005). As these phenomena of nonword repetition have been viewed as offering support for the PSTM account, they are particularly relevant to the present investigation. If the model exhibited them, it would instantiate some key behaviors that motivate the PSTM account.

The three bars at the left of Fig. 5(a) display repetition accuracy for two-, three-, and four-syllable nonwords from Gupta et al. (2005), which illustrate the basic phenomenon of nonword repetition accuracy decreasing with nonword length, for adults. The Figure also displays nonword repetition accuracy data from Simulation 2 at an adult level of 8000 epochs, now broken down by nonword length. (Note that these behavioral and simulation data are identical to those shown collapsed across nonword length in Fig. 4.) As can be seen, the model's performance exhibits a clear effect of nonword length, and furthermore, matches the behavioral data quite well. The main effect of length was significant in both the human and simulation data, $p < .005$ in each case.

We also examined the effect of nonword length at an earlier developmental stage. As discussed in Simulation 1, the model showed a clear trajectory of phonological vocabulary development, asymptoting in the 6000–8000 epoch range, which we therefore termed the adult range. If this trajectory is plausible, then the model's nonword repetition performance at earlier points would have at least some correspondence with the behavioral data for younger age groups. We did not expect this correspondence to be very systematic, and merely expected there to be points of contact such that at some younger "ages" of the model, nonword repetition performance would correspond with that of children.

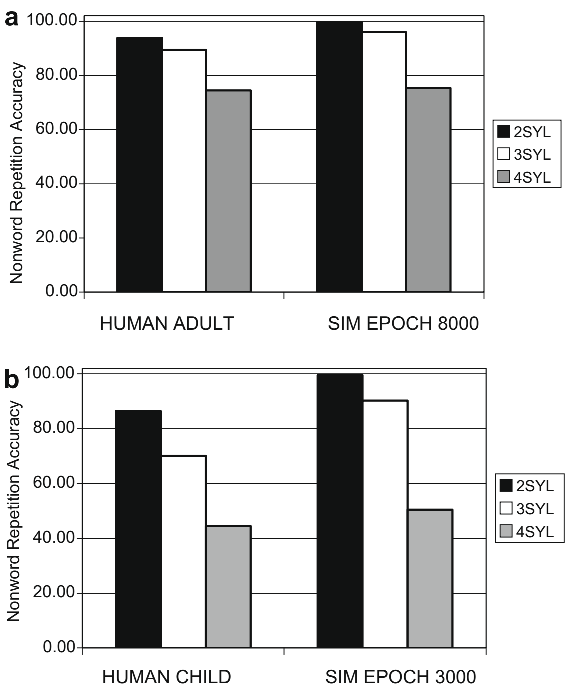

Fig. 5. Effect of nonword length on nonword repetition accuracy: data and simulations. (a) Human adults (Gupta et al., 2005) and model at 8000 epochs. (b) Human children (Gupta et al., in preparation) and model at 3000 epochs.

However, to provide some informal sense of how the model's trajectory might map onto human age, suppose we assumed a linear relationship between epochs in the model and age. We note again that we do not expect or claim that such a systematic mapping actually holds, and thus we introduce it only for illustrative purposes. Under this linear assumption, if we consider epoch 6000 (the start of the adult range in the model) to correspond loosely with an age of 18 years, then epoch 3000 would correspond loosely with an age of 9 years. With this loose analogy in mind, we compared the model's nonword repetition at 3000 epochs with that of a group of children aged 6–11 with a mean age of 9.02 years (Gupta, Abbs, Tomblin, & Lipinski, in preparation). The stimuli presented to the children came from the same corpus as those presented to the model. The three bars at the left of Fig. 5(b) display repetition accuracy for two-, three-, and four-syllable nonwords in the children, while the three bars at right display the model's performance from Simulation 2 at 3000 epochs, broken down by nonword length. Here, too, the model displays a clear effect of nonword length, and also matches the behavioral data reasonably well. Again, the effect of length was significant in both the human and simulation data, $p < .005$ in each case.

We further examined serial position effects within nonwords. Gupta (2005) and Gupta et al. (2005) demonstrated that repetition of nonwords was subject to syllable-wise serial position functions, and this finding has also been replicated in other age groups and languages (e.g., Archibald & Gathercole, 2007; Gupta et al., in preparation; Santos, Bueno, & Gathercole, 2006). Fig. 6(a) displays serial position functions obtained in repetition of two-, three-, and four-syllable nonwords from adults (Gupta et al., 2005). Fig. 6(b) shows accuracy by serial position for nonword repetition at 8000 epochs from Simulation 2. (The functions in Fig. 6a and b are derived from the behavioral and simulation results plotted by nonword length in Fig. 5a.) As can be seen, the model corresponds quite well with the behavioral data. In the behavioral data, classic primacy and recency are seen most clearly for four-syllable nonwords, and this is the case for the model as well. The model also displays the lack of primacy as well as lack of recency seen in the behavioral data for three-syllable nonwords. The overall levels of accuracy are quite similar to those in the behavioral data.

Fig. 7(a) shows serial position functions for 6- to 11-year-old children (Gupta et al., in preparation), and Fig. 7(b) serial position functions from the model at 3000 epochs, both corresponding to the data plotted by nonword length in Fig. 5(b). Here, too, the model captures the behavioral data fairly well, showing clear primacy and recency for four-syllable nonwords, and primacy and a lack of recency for three-syllable nonwords. However, the model does not capture the negative recency for three-syllable nonwords, and it shows slight primacy for two-syllable nonwords, where there is none in the behavioral data. Overall levels of performance are in a roughly similar range to the behavioral data.

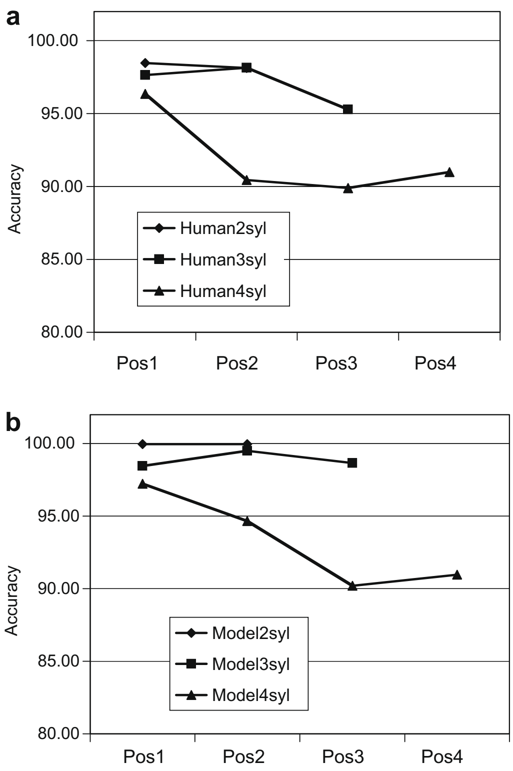

Fig. 6. Nonword repetition accuracy by syllable serial position, for (a) adults (Gupta et al., 2005), (b) simulation at 8000 epochs.

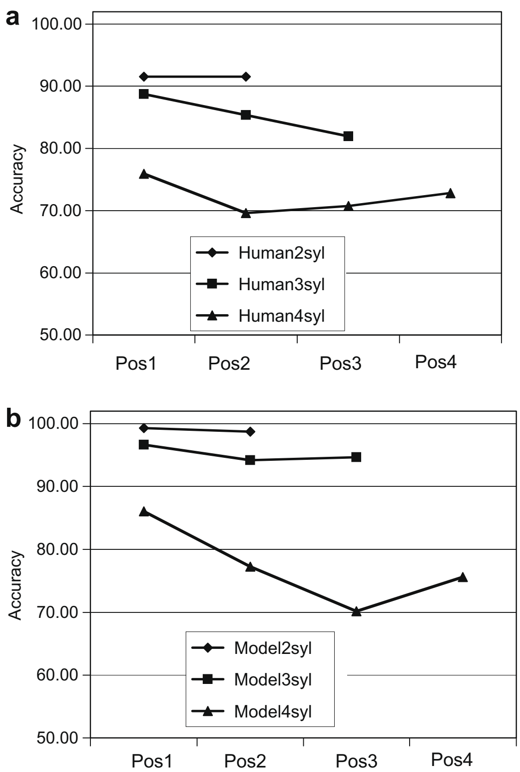

Fig. 7. Nonword repetition accuracy by syllable serial position, for (a) children (Gupta et al., in preparation), (b) simulation at 3000 epochs.

In summary, then, the model's nonword repetition exhibits behavioral effects that in humans have been attributed to PSTM, thus offering a computational account of these effects, and establishing that the model simulates key phenomena motivating one of the two accounts. We note once again that these results involved no manipulation of the model, and are merely finer-grained analyses of the model's nonword repetition performance reported in Simulation 2 – that is, they fall out of the model.

Effects of phonotactic probability: Simulation 3

As noted in the Introduction, the findings that nonword repetition accuracy is influenced by lexical/phonological knowledge (e.g., Gathercole, 1995) and that these effects change with vocabulary size (e.g., Munson et al., 2005) have formed an important part of the rationale for the linguistic account. Gathercole (1995) found, for example, that 4- and 5-year-old childrens' nonword repetition accuracy was greater for highly wordlike nonwords than for less wordlike nonwords, exemplifying an effect of phonological knowledge on nonword repetition performance. As wordlikeness is highly correlated with phonotactic probability (e.g., Munson, 2001), these results have also been interpreted as showing an effect of phonotactic probability (Gathercole, 2006). Munson et al. (2005) further found that the effect of phonotactic probability on childrens' nonword repetition accuracy decreased as a function of vocabulary size, also suggesting that phonological knowledge impacts nonword repetition performance.

To examine whether the phonological knowledge encoded in the model gives rise to such effects, we created 100 word forms of each of four types: two-syllable high phonotactic probability, two-syllable low-phonotactic probability, three-syllable high phonotactic probability, and three-syllable low-phonotactic probability. None of these stimuli was in the vocabulary to which the model was exposed as part of learning in Simulation 1. Positional segment (i.e., phone) probabilities and biphone probabilities were calculated using the method described by Vitevitch and Luce (2004). Phone and biphone probabilities were significantly lower for the low probability stimuli than for the high probability stimuli at each of the two syllable lengths (two-syllable low vs. high mean phone probability: .199, .264; two-syllable low vs. high mean biphone probability: .005, .032; three-syllable low vs. high mean phone probability: .194, .308; three-syllable low vs. high mean biphone probability: .005, .035; $p < .0005$ for all low vs. high comparisons). At a number of points during the model's vocabulary development as described in Simulation 1, we tested its performance in nonword repetition for each of these high and low-phonotactic probability stimuli, using the procedure of Simulation 2, and the model of Simulation 1.

Fig. 8(a) plots the results reported by Gathercole (1995) for 4- and 5-year-olds, while Fig. 8(b–d) plot the model's

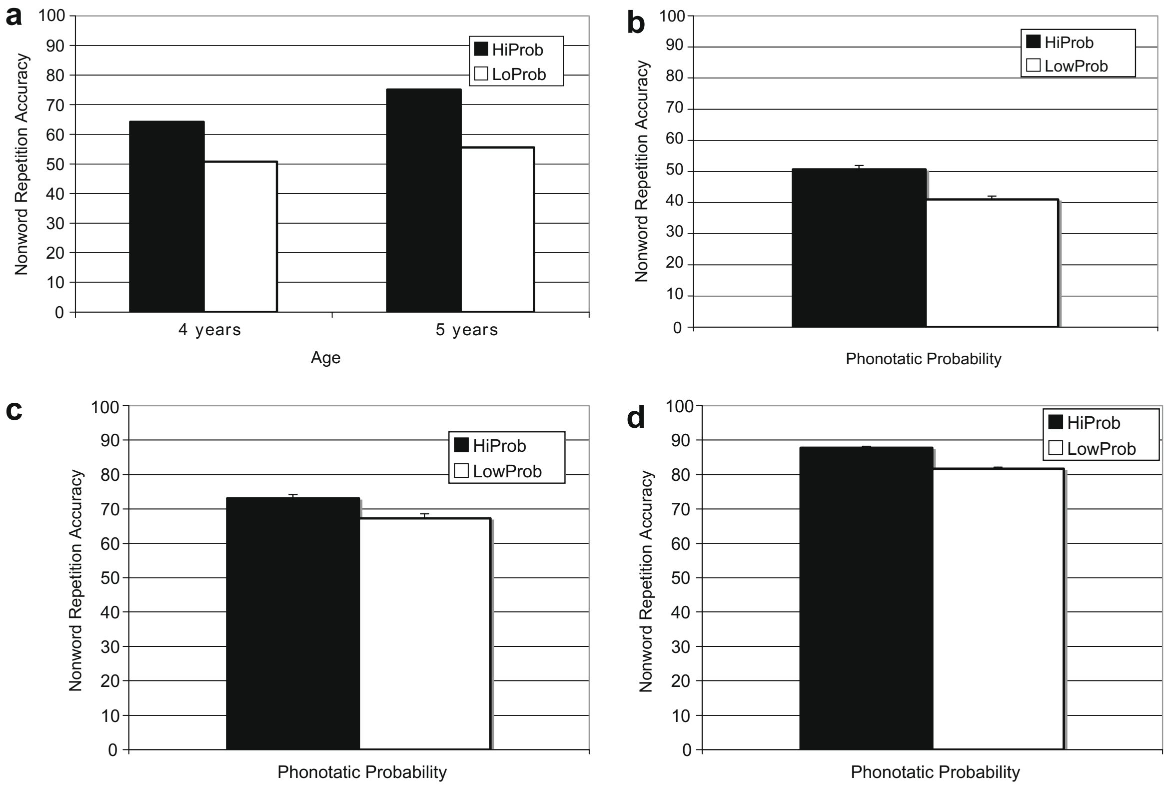

Fig. 8. Effect of phonotactic probability on nonword repetition accuracy. (a) Effect in 4- and 5-year-old children, re-drawn from data in Gathercole (1995). (b–d) Effect in model, at three "ages": epochs 500, 2000, and 8000.

nonword repetition performance for the high and low-phonotactic probability nonwords in Simulation 3 at three points chosen to correspond roughly to a "young child" age (500 epochs), an "older child" age (2000 epochs) and an "adult" age (8000 epochs). We did not attempt to set these epochs into any more systematic correspondence with age. At each of these points, the model's nonword repetition was significantly more accurate for the high phonotactic probability than the low-phonotactic probability words ($p < .001$ at each simulated age).[2] As can be seen in the figure, the size of the effect was smaller in the model than in the empirical data; nevertheless, the model did show a clear effect of phonotactic probability. The occurrence of these phonotactic probability effects in the model falls out of the very manner of its operation and learning: phones and combinations of phones that occur more frequently in the vocabulary have been processed more frequently by the model, leading to better performance in generalizing to new stimuli (i.e., nonwords) that incorporate them. This is very much the same explanation that has typically been offered for the effects of phonotactic probability in human nonword repetition (e.g., Edwards et al., 2004); however, the present model provides a concrete instantiation of and mechanistic basis for such an explanation.

Munson et al. (2005) reported a diminution of the effects of phonotactic probability in typically developing children over the age range 60–160 months such that an interaction between phonotactic probability and age was found (Fig. 3, page 1043, Munson et al., 2005). To analyze whether such an effect existed in the model, we conducted a 2-way ANOVA comparing the model's performance on high and low-phonotactic probability nonwords at the vocabulary sizes corresponding to epochs 500 and 2000, the two "child" ages from Fig. 8. The ANOVA revealed a significant interaction ($p < .0005$), plotted in Fig. 9. In the model, as in Munson et al. (2005) results, the effect of phonotactic probability diminishes with age. The reason for this in the model is that very early in phonological vocabulary learning, weight adjustment in the model has been dominated by higher probability phones and phone combinations, because of their relatively high frequency of occurrence, so that accuracy at 500 epochs is substantially greater for stimuli containing higher probability than lower probability elements. By later in phonological vocabulary learning, higher probability elements no longer

[2] Assessing statistical significance required that multiple observations of nonword repetition be obtained from the model, at each "age". This was achieved by testing the model at each of the 25 epochs immediately preceding a chosen age value. Thus, for instance, nonword repetition performance for epoch 2000 was assessed by measuring nonword repetition performance at epochs 1975 through 2000. This can be thought of as analogous to sampling 25 individuals of about (but not exactly) the same "age".

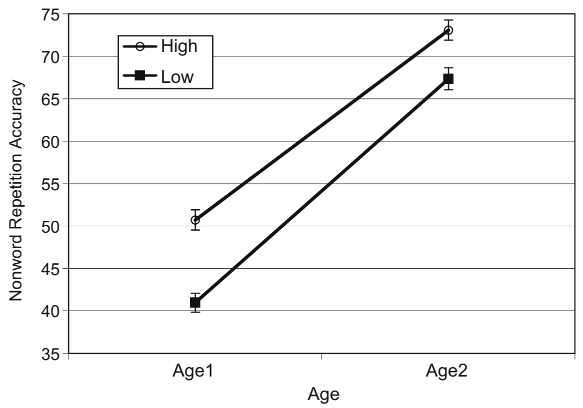

Fig. 9. Interacting effects of age and phonotactic probability on nonword repetition: simulation.

generate as much error (their accuracy is closer to ceiling), whereas lower frequency phones and phone combinations continue to generate substantial error. Error-driven weight adjustment therefore now occurs more in response to the lower frequency elements, so that their accuracy begins to "catch up", e.g., at 2000 epochs. The model thus offers an explanation of this phenomenon as well.

The interaction noted above indicates that the difference between repetition accuracy for high- versus low-phonotactic probability nonwords (i.e., the phonotactic probability effect) decreased with age. Munson et al. (2005) also found, however, that this effect (which they termed the *frequency effect*) decreased as a function of vocabulary size as gauged by a standardized vocabulary measure, the Peabody Picture Vocabulary Test, or PPVT (Fig. 2, p. 1043, Munson et al., 2005). Munson et al. (2005) therefore examined the relative contributions of age and vocabulary size to the phonotactic probability effect by conducting a hierarchical regression in which first age and then the vocabulary measure were entered as predictors. The results of this analysis, summarized in the upper panel of Table 2, indicated that age and vocabulary size made independent contributions to the phonotactic probability effect (Table 4, p. 1042, Munson et al., 2005). However, it was only the effect of vocabulary size that was negative (as indicated by the negative beta weight in Table 2), while the effect of age was positive, although not significantly so. That is, it was increases in vocabulary size rather than age that led to a decrease in the phonotactic probability effect. Also, as indicated by the ΔR^2 values in Table 2, vocabulary size accounted for a greater proportion of variance in the phonotactic probability effect than did age. To examine the effects of age and vocabulary size in the model, we conducted a hierarchical regression of the phonotactic probability effect between epochs 500 and 8000, incorporating first epochs (the model's analogue of age) and then phonological vocabulary size as the predictors, as in Munson et al. (2005). The results are shown in the lower panel of Table 2. Although age is a significant predictor in the model, thus differing from the empirical results, the model in other key respects shows the same pattern as reported by Munson et al. (2005): a negative beta weight for vocabulary size, but not for age, and a higher ΔR^2 for vocabulary size than for age. Thus in the model as in the behavioral data, the longitudinal decrease in the phonotactic probability effect is driven more strongly by increasing vocabulary size than by age.

Understanding the model's functioning

The simulations thus far establish that the model learns a phonological vocabulary, and that its nonword repetition performance exhibits overall levels of accuracy similar to those of human adults and children, shows nonword length effects and serial position effects of the kind that support the PSTM account, and shows phonotactic probability effects of the kind that support the linguistic account. What underlies these behaviors and effects in the model? And what, if anything, does PSTM have to do with them?

Serial order processing in the model

We can begin by noting that the task performed by the model in all of these behaviors is the production of word forms as serially ordered sequences. Let us examine how the model achieves this serially ordered behavior. (See also Botvinick & Plaut, 2006, who present extensive analyses of their model's functioning. The principles of sequential operation are largely similar in the two models, even though the Botvinick & Plaut, 2006 model was applied to sequences of letters rather than syllables, and the focus was not on learning specific sequences.)

When input of the sequence of syllables representing a particular word form has ended, the model is left with the task of replicating that particular sequence, in the absence of any external input other than the "Recall" signal (which is identical for all word forms, and hence provides no useful information about the content of any particular word form). Two things are needed to solve this computational

Table 2
Hierarchical regression analyses of the contributions of age and vocabulary size to the phonotactic probability effect: comparison of behavioral data (Munson et al., 2005) and simulation.

Data	Order of entry	Variable entered	ΔR^2	Significance	*Beta* weight	Significance
Munson et al. (2005)	1	Age	.081	$p < .05$	.05	*ns*
	2	PPVT vocab	.148	$p < .01$	−.53	$p < .01$
Simulation	1	Epoch	.243	$p < .005$	.416	$p < .05$
	2	Vocab size	.429	$p < .001$	−1.121	$p < .001$

problem. First, the activation pattern over the hidden units at the end of input must encode information about the entire word form, including its serial order. Second, this hidden layer activation pattern must change during recall, such that at each time step, it provides the basis for the model to produce the correct syllable at the output.

To clarify the second point: If the hidden layer activation pattern remained the same throughout all time steps of recall, then, even if it constituted a representation of the entire word form, the model's output would remain a static pattern, rather than a sequence of appropriate syllables. In other words, for different outputs to be produced from one time step to the next, the hidden layer activation must change over these time steps. For the hidden layer activation pattern to change from one time step of recall to the next, the model's input must change. Input from the "Recall" signal, however, is unchanging throughout. The change in the model's input therefore has to come from elsewhere – from the context information. Thus the model's ability to modify its hidden layer activation appropriately during recall depends critically on the recurrent connections on the hidden layer, which provide the context information (which is derived from the hidden layer's own previous state). Without these connections, and in the presence of the unchanging and uninformative "Recall" input, no change would occur in the model's input and hence hidden layer activation pattern, and hence output. The model would just keep producing the same output, time step after time step.

The model's context information, provided by its hidden layer recurrent connections, is thus critically necessary for its serially ordered production of word forms. This speaks to the first question we raised in Introduction: (1) What are the computational requirements of serial order production in the nonword repetition task? Given that nonword repetition is a serially ordered word form production task, it follows that having and using context information is a computational requirement for nonword repetition.

What is phonological short-term memory?

The second question we raised in Introduction pertained to whether short-term memory has any relationship to these computational requirements. To examine this question, let us further examine the nature of the context information.

Because the context information is provided by recycling one of the network's own previous activation patterns, or states, it is *information about the past*-information about what past state(s) the system has been in, during production of the sequence. But information about past states is nothing if not *memory* information. Furthermore, because the context information is reset from one sequence to the next, it pertains only to the sequence *currently* being processed, and so it is *short-term* memory information. (It is also constantly overwritten even during the processing of a particular sequence, again making it short-term memory information.) Thus in our model, and in recurrent neural networks more generally, short-term memory information *is* context information; and as discussed above, context information is an inherent computational requirement for serially ordered production of sequences.[3] From this it follows that in a recurrent neural network, short-term memory information is computationally necessary for the serially ordered behavior of nonword repetition. This addresses the second question we raised above, regarding the relationship of short-term memory to nonword repetition.

The hidden layer activation pattern at the end of input presentation of a word form is the model's encoding of the entire word form, as noted previously, and this becomes context information at the next time step, and thus the model's "STM encoding" of the word form. Is this a phonological encoding? If so, then the model's STM would also be its *phonological* STM. Given that the model employs input and output representations that are distributed and phonologically structured, and given that its hidden layer representations are also distributed, we would certainly expect the hidden layer encodings of word forms, and hence the STM encodings, to be phonologically structured. Nevertheless, to confirm this, we created one set of nonwords that rhymed with each other, and one set of nonwords that were dissimilar to each other. We verified that, as would be expected, the input representations of the rhyming word forms were more similar to each other than the input representations of the dissimilar word forms were to each other $t(189) > 32.0$, $p < .0001$. Thus the input layer vector representations of the two sets of nonwords reflected the phonological similarity relations that the nonwords were designed to incorporate. The question of interest was whether the model's STM encodings of these nonwords would reflect these phonological similarity relations. To examine this, we presented all the nonwords to the model, the procedure being exactly as in Simulations 2 and 3. For each nonword, we recorded the model's STM encoding of it – the context information present in the model at the beginning of recall of that nonword. Within-set similarity was greater for the STM encodings of the rhyming nonwords than the dissimilar nonwords, $t(189) > 38.0$, $p < .0001$. Thus the model's STM encodings reflect the phonological similarity of the input, confirming

[3] The availability of context information is in fact a computational requirement for the production of serially ordered sequences, not only in neural networks, but in general. The many mathematically and/or computationally specified accounts of various aspects of serially ordered behavior differ in many respects (e.g., Botvinick & Plaut, 2006; Brown et al., 2000; Burgess & Hitch, 1992, 1999; Christiansen et al., 1998; Dell et al., 1993; Elman, 1990; Grossberg, 1978, 1986; Gupta, 1996, 2009; Gupta & Cohen, 2002; Hartley & Houghton, 1996; Houghton, 1990; Jordan, 1986; Page & Norris, 1998), but all converge on one notion: in all of these accounts, producing a serially ordered sequence has the computational requirement that *context* information must be available – that is, information that can serve to distinguish successive states of the system from each other, during production of the sequence. It should be noted that our use of the term "context" is not meant to be equated with any one particular mechanism that has been proposed in these various computational models – for instance, a time-varying *context signal* (e.g., Brown et al., 2000; Burgess & Hitch, 1992, 1999; Gupta, 1996; Hartley & Houghton, 1996) or updating of a *context layer* (e.g., Elman, 1990) as in the present model, or changes in the pool of activated items as a result of items' post-recall suppression (Page & Norris, 1998), all of which serve to distinguish successive states of the system from each other. Our use of the term is meant to abstract away from particular instantiations, and refers to an abstract functionality that is needed for serial order processing.

that the model's short-term memory is, indeed, *phonological short-term memory*. This elaborates our answer to the second question we raised in Introduction, indicating that *PSTM* does play a necessary role in the repetition of novel word forms.

Two further points are important here. Because the PSTM information available in the context layer is transmitted to the hidden layer across weighted connections, the PSTM or context information that is actually received by the hidden layer is colored by those connection weights. But these weights are long-term weights, set as a result of the network's entire experience – like any long-term weights in a neural network, they incorporate its long-term knowledge (in this case, phonological knowledge). From this it follows that in the model, phonological short-term memory information is influenced by long-term phonological knowledge. Said differently, there is no such thing as a "phonological short-term memory" that is independent of a "long-term memory".

In addition, there is no one structure or process in a recurrent network that *is* "phonological short-term memory" – rather, there is PSTM functionality, which arises from the joint operation of many structures and processes, such as the weights of the recurrent connections, other connection weights in the system, and the mechanisms of activation propagation; and this functionality is inherently intertwined with long-term knowledge in the system. For these reasons, we will henceforth use the more neutral term *phonological short-term memory functionality*, defining it as the functionality of "having phonological information about the recent past". This term is intended to avoid the connotation that "phonological short-term memory" is a process or store, which becomes problematic if "phonological short-term memory" cannot be identified with any particular structure or process, and cannot be cleanly separated from "long-term memory". Our definition of the term is consistent with the possibility that the information about the recent past may itself be colored by long-term knowledge, i.e., by the not-so-recent past.

Length effects and serial position effects

The effect of length on nonword repetition accuracy is a direct consequence of the nature of the representations that the model constructs for word forms. As discussed above, the hidden layer representations that develop by the end of input of a word form are what guide its subsequent recall. Botvinick and Plaut (2006) showed that a model of this type can be thought of as "unfolding" this hidden layer activation pattern in time, in order to produce the correct sequence element at the correct time step, and that the greater the number of sequence elements (in the present model, syllables), the more difficult it is to create an encoding that maintains sufficient discrimination between those elements. (Note that this account falls into the broad category of "interference" accounts, and does not incorporate trace decay.) Length effects thus fall out of the representations that the model constructs, with nothing being added to the model to achieve them. Performance differences at the "child" and "adult" ages also fall out of the model's manner of operation (as the result of continuous weight change throughout simulated development), with no manipulation of model parameters.

The explanation of serial position effects is also related to the discriminability of components of the hidden layer encodings of word forms. For a *given* number of syllables, maintaining discriminability is easiest for syllables that are distinguished by being at a boundary (either the beginning or the end of a word form – an "edge" effect), and more difficult for syllables in the middle of the word form, thus giving rise to serial position effects. Again, the explanation is exactly as in the Botvinick and Plaut (2006) model of immediate serial recall of lists of letters.[4]

Effects of phonological knowledge

As we have seen, during learning, the model develops internal representations of the word forms it is exposed to, and learns how to unpack these internal representations in sequence. That is, it acquires knowledge about phonological forms, and about the transition structure within these forms. Moreover, because this knowledge must be developed for *all* word forms in the *same* set of connection weights, the connection weights must encode knowledge of the entire vocabulary. This knowledge in turn affects the processing of word forms, because the connection weights are crucial determinants of all processing in the model. It is for this reason that effects of phonological knowledge arise in the model.

It is also worth noting that the model's functioning can be seen as providing a mechanistic instantiation of the construct of redintegration referred to earlier, according to which long-term memory plays a role in reconstruction of the short-term memory trace. As we have discussed, the context information available at the beginning of recall of a word form can be thought of as the STM trace of the presented word form; its use in production of the word form is therefore analogous to reconstruction of the to-be-recalled stimulus from an STM trace. As we have also discussed, the "spelling out" of this hidden layer encoding is critically dependent on the model's recurrent connections, whose weights encode long-term knowledge. The influence of long-term knowledge on the model's reconstruction is thus functionally equivalent to redintegration. (See also Botvinick & Plaut, 2006 for discussion of this point for the domain of immediate serial list recall.)

The causal roles of PSTM functionality and vocabulary size/knowledge

We have discussed how PSTM functionality in the model is critical to its nonword repetition ability. We are now in a position to directly address the key theoretical questions on which the PSTM and linguistic accounts differ: Does PSTM functionality in the model causally affect its phonological vocabulary learning? And, does phonological vocabulary growth in the model cause improvement in

[4] Such "edge effects" also constitute a basis for primacy and recency effects in a number of other computational models of serial order production (e.g. Burgess & Hitch, 1992; Gupta, 1996, 2009).

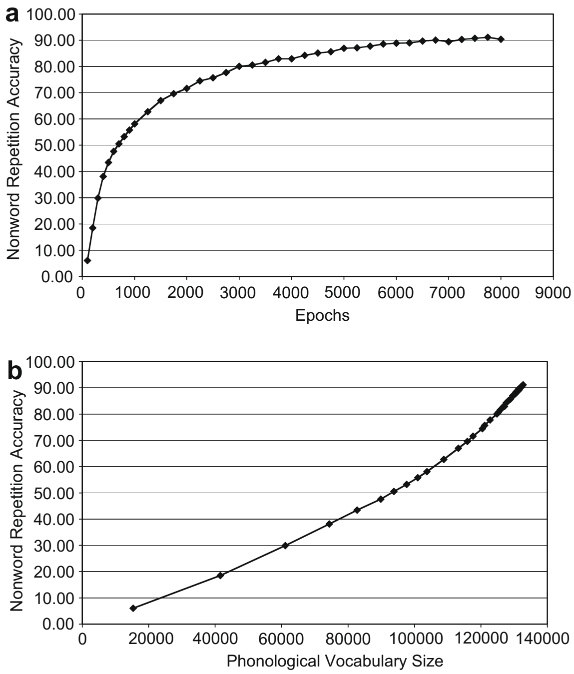

Fig. 10. Repetition accuracy for nonwords across development (simulation). (a) As a function of epochs. (b) As a function of phonological vocabulary size.

nonword repetition ability, independent of any change in parameters that affect PSTM functionality? We begin by addressing the second question.

The effect of vocabulary size/knowledge on nonword repetition accuracy: Simulation 4

A central hypothesis of the linguistic account is that nonword repetition performance is influenced primarily by the level of phonological vocabulary size/knowledge available to the individual, and, accordingly, that the association observed between nonword repetition ability and vocabulary size is largely a result of this influence. To test this hypothesis, we examined whether the model's nonword repetition performance would improve as a function of its phonological vocabulary/knowledge, without any explicit adjustment of model parameters that could affect PSTM functionality. To do this, we used the same two-, three-, and four-syllable nonwords as in Simulation 2, again testing the model's repetition of these stimuli. Now, however, we did this at numerous epochs during the model's development described in Simulation 1. As in the previous simulations of nonword repetition, the procedure consisted of providing the nonword as input to the model, one syllable at a time, and determining the accuracy of the model's repetition. Connection weights were not adjusted at any point, so that at each epoch, the model's nonword repetition was based solely on generalization from its existing phonological vocabulary/knowledge at that epoch, rather than on any prior learning of the specific nonwords presented. All parameters that might affect PSTM functionality (and indeed, all parameters of the model) remained unchanged throughout the epochs of development in Simulation 1, as in all simulations we have reported.

Fig. 10(a) plots the model's nonword repetition performance across the 8000 epochs of phonological vocabulary learning described in Simulation 1, collapsed across nonword length. As can be seen, the model's nonword repetition accuracy exhibits a clear developmental trend as a function of epochs (i.e., simulated age). That is, nonword repetition accuracy improves as a function of the model's increasing phonological knowledge, with no adjustment having been made to PSTM functionality. Fig. 10(b) shows even more clearly that this improvement is a function of phonological knowledge, plotting nonword repetition accuracy against phonological vocabulary size rather than epochs. Here the trend is linear or even exponential. Nonword repetition accuracy is strongly influenced by phonological vocabulary size. Thus the model supports this tenet of the linguistic account.

The causal effects of PSTM functionality: Simulations 5 and 6

To examine the causal effect of PSTM functionality on phonological vocabulary learning, we manipulated the effectiveness of that functionality. As explained earlier, the connections from the hidden layer to the context layer provide for a copy of the hidden layer's activation at the end of each time step, and all these connections have the same fixed weight. *What* that fixed weight is, however, can be manipulated. In all simulations discussed so far, these weights were set at 1.0, as is typical in simple recurrent networks. This amounts to a completely veridical or undecayed transmission of the hidden layer's activation pattern to the context layer. However, different values of this fixed connection weight would be expected to affect the veridicality of the copy ("PSTM maintenance strength") and hence PSTM functionality. We therefore implemented a *pmaint* parameter to vary the value of the fixed weight of these connections and thus PSTM maintenance strength and PSTM functionality.[5] If such variation affected the model's phonological vocabulary learning, it would evidence a causal effect of PSTM functionality. In addition, to concretize earlier discussion of the need for PSTM functionality in non-

[5] Note that the value of the fixed connection weights can equally well be thought of as expressing either activation maintenance or activation decay (decay and activation maintenance are the complements of each other), and the parameter can therefore be thought of as controlling either activation maintenance or trace decay. The notion of decay has not previously been incorporated in distributed sequential learning models (e.g. Botvinick & Plaut, 2006; Chang, Dell, & Bock, 2006; Elman, 1990; Gupta & Cohen, 2002; Jordan, 1986). It does, however play an important role in many other neural network models (e.g., McClelland & Rumelhart, 1981), that typically make greater use of localist representations. In such models it is well established that recurrence-based short-term memory functionality depends on both decay and connection weight strength. However, in such models, the recurrence is not employed for sequential processing: Although the information about the past that is provided by the recurrence-based STM functionality influences subsequent time steps of processing, so that the system exhibits dynamical behavior in settling to a stable activation state, the sequence of states does not replicate a specific target sequence. Incorporation of decay into the present model can thus be seen as an extension of ideas from such recurrent but non-sequential models to recurrent sequential learning models.

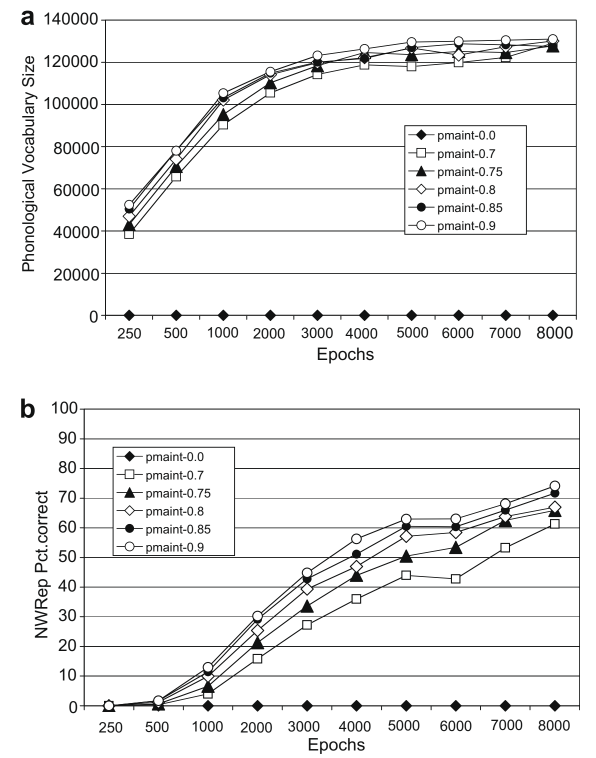

Fig. 11. Simulations with varying or abolished PSTM functionality. (a) Phonological vocabulary learning. (b) Nonword repetition.

word repetition, we also examined the effect of such variation on nonword repetition accuracy.

As noted above, the value of *pmaint* was 1.0 in all simulations thus far. We now used six lower values: .90, .95, .80, .75, .70, and 0. The last of these values was equivalent to a complete abolition of PSTM functionality. We constructed six different networks, each incorporating one of these *pmaint* values, and each of these networks was used to simulate phonological vocabulary learning, as in Simulation 1. At various epochs during this learning, and when each network had reached 8000 epochs of learning, we recorded its vocabulary size. We also tested its nonword repetition (as in Simulation 2) of 100 four-syllable nonwords at each of these epochs. The question of interest was whether the variation in PSTM functionality would result in systematic differences in the developmental trajectory of both phonological vocabulary learning and nonword repetition.

Fig. 11(a) shows the results of Simulation 5, which examined the developmental trajectory of phonological vocabulary learning under variation in *pmaint*. The six different curves indicate the learning trajectories for the six *pmaint* values. The plots indicate clearly that phonological vocabulary learning in the model is systematically affected in a graded manner by graded variation in PSTM functionality, and completely eliminated by abolition of PSTM functionality. Fig. 11(b) shows the results of Simulation 6, which examined the model's nonword repetition accuracy at these same epochs, for each of the six *pmaint* values. The Figure shows clearly that accuracy is affected in a graded manner by graded variation in PSTM functionality and that nonword repetition ability is eliminated by abolition of PSTM functionality. Together, these results indicate that PSTM functionality does causally affect phonological vocabulary learning, and illustrate how it affects nonword repetition ability. As a further demonstration, we conducted an additional simulation using the model of Simulation 1 at 8000 epochs, i.e., after the model had developed a normal "adult" phonological vocabulary with the *pmaint* parameter set at 1.0. At 8000 epochs, we reduced the value of *pmaint* to 0, thereby eliminating PSTM functionality without altering any other connection weights or any other aspect of the model. Even though the model retained all other connection weights and thus its long-term phonological knowledge, its nonword repetition ability was abolished. All of these findings provide clear support for the PSTM account.

Recapitulation: Reconciling the PSTM and linguistic accounts

The results of Simulations 4–6 are clear in establishing three things. First, PSTM functionality is a critical and causal determinant of phonological vocabulary learning – without PSTM functionality, there is no vocabulary learning/growth (Simulation 5), and such learning is also affected in a graded manner by variation in PSTM functionality. Second, as our earlier theoretical analysis indicated, and as Simulation 6 illustrated, PSTM functionality is a critical and causal determinant of nonword repetition ability. Both of these determinations are in keeping with the postulates of the PSTM account. Third, nonword repetition ability increases as a function of increasing vocabulary size/knowledge, even when no aspect of the model's PSTM functionality has been explicitly manipulated (Simulation 4). This is in keeping with the linguistic account. Together, these results clearly indicate that there is no opposition between the key tenets of both the PSTM and linguistic accounts. Given the model's credibility, as indicated by its ability to capture a range of behavioral phenomena (including findings that have been adduced as divergent evidence for each of the two accounts), there is considerable reason to suppose that these causal findings from the model are also true of human cognition.

Fig. 12(a) depicts the relationships revealed by Simulations 4–6. It depicts the causal links from PSTM functionality to nonword repetition ability and to phonological vocabulary learning, as in the depiction of the PSTM account in Fig. 1(a). (Note, however, that the construct of *PSTM* in Fig. 1 has been replaced by the construct of *PSTM functionality*.) Fig. 12(a) also depicts the causal link from phonological vocabulary learning to phonological vocabulary size/knowledge (a logical necessity, as noted in Introduction, and incorporated as a background assumption in both the PSTM and linguistic accounts in Fig. 1). In addition, Fig. 12(a) shows the link evidenced by Simulation 4, from phonological vocabulary size/knowledge to nonword repetition ability (as in the depiction of the linguistic account in Fig. 1b). The model thus provides a clear reconciliation of the PSTM and linguistic accounts.

The pattern of causality is, however, more complex than this. The complexity has to do with the nature of PSTM functionality. As discussed in Introduction and under "What is

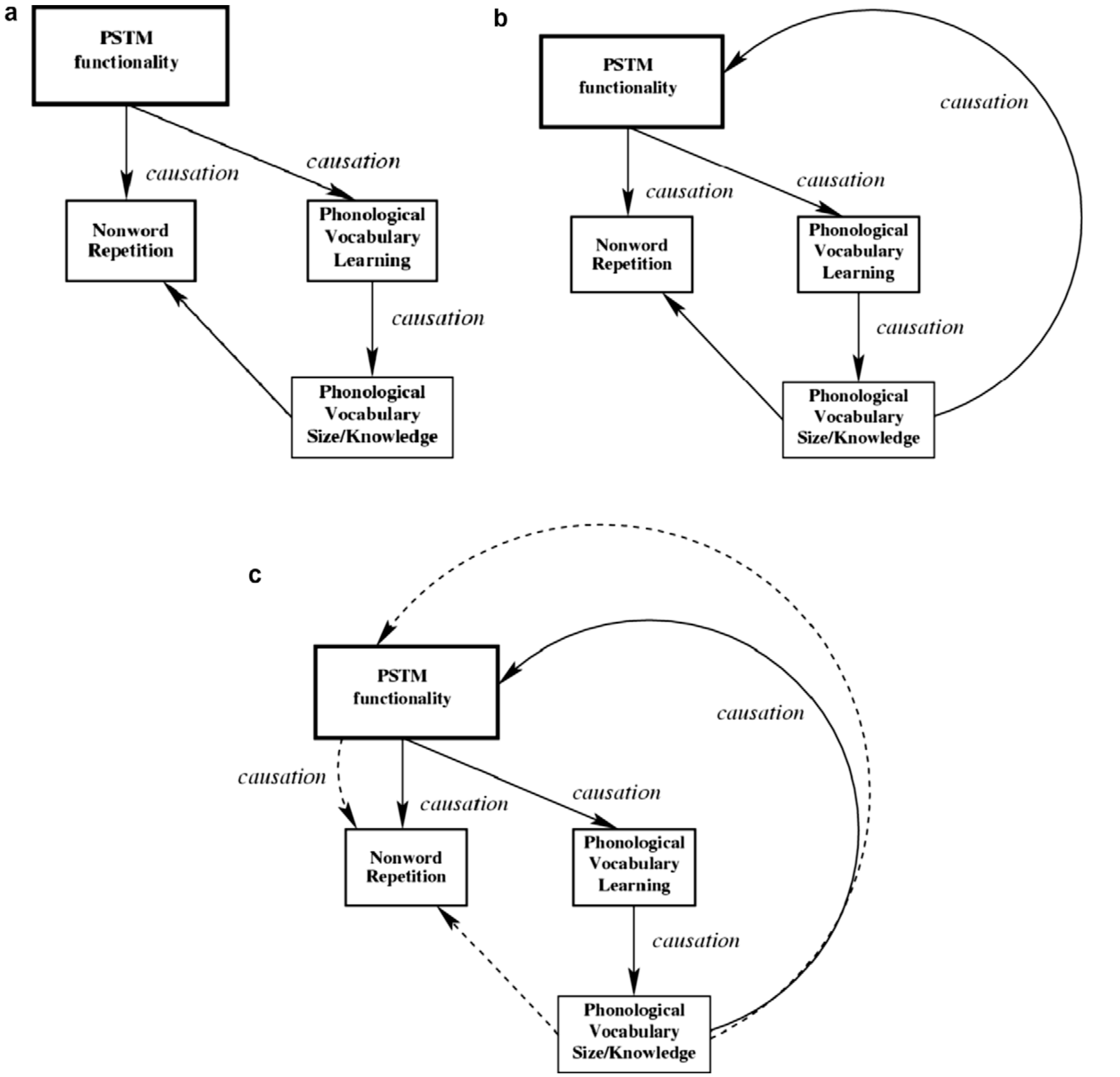

Fig. 12. Patterns of causality revealed by the simulations. (a) Initial formulation. (b) Revised formulation, reflecting the causal effect of phonological vocabulary size on PSTM functionality. (c) Third formulation, showing two possible paths for causal effect of phonological vocabulary size on nonword repetition performance.

Phonological Short-Term Memory?", PSTM functionality is itself dependent on long-term knowledge in the system. Thus greater long-term phonological knowledge causes greater PSTM functionality – this is an inherent consequence of the way PSTM functionality works in the model. Fig. 12(b) introduces this path of causality, indicating it by the new causal link from phonological vocabulary size/knowledge to PSTM functionality. It is worth emphasizing that this does not alter the finding that PSTM functionality is causally critical for both phonological vocabulary learning and for nonword repetition ability. Phonological vocabulary learning simply cannot occur without the serial ordering capability provided by PSTM functionality, no matter what the vocabulary size. Similarly, nonword repetition (or word repetition for that matter) is simply not possible without PSTM functionality, no matter what the vocabulary size. However, PSTM functionality itself is causally affected by phonological vocabulary size/knowledge. The relationship between PSTM functionality and phonological vocabulary size/knowledge is thus a bidirectional one, making this an interactive (and thus more complex) system.

In addition to making for a more complex picture in general, this has specific implications for the association between phonological vocabulary size/knowledge and nonword repetition. Notice that Fig. 12(a) and (b) do not label this link as causal. This is because there are actually two possible causal paths that might underlie this association. It could arise in part from a *direct* causal impact of greater phonological vocabulary size/knowledge on nonword repetition ability, as the arrow in Fig. 12(a) and (b) implies (but does not explicitly depict). However, it is certain from the manner of the model's operation that there is an *indirect* path for causality whereby phonological vocabulary size/knowledge causally affects PSTM functionality which then causally affects nonword repetition ability. The direct path might or might not actually be a causal path. The indirect path must necessarily be a causal path, but might or might not account for a significant part of the association. Thus the association might be explained by either or both types of causality. Fig. 12(c) revises Fig. 12(b) to depict these possibilities, showing two dashed lines that represent the two possible causal paths underlying the association between phonological vocabulary size/knowledge and nonword repetition ability. (Causality from nonword repetition ability to phonological vocabulary size/knowledge can be ruled out – the model makes it clear that performing nonword repetition does not lead to any change in knowledge.) Furthermore, because phonological vocabulary size/knowledge causally affects PSTM functionality, which causally affects phonological vocabulary learning, this means there is a feedback loop from PSTM functionality to vocabulary learning to phonological vocabulary size/knowledge to PSTM functionality to phonological vocabulary learning, as can be seen in Fig. 12(b) and (c). Thus even while supporting the postulate of the linguistic account, the model reveals greater underlying complexity.

General discussion

The present work offers a computational account of phonological vocabulary learning and nonword repetition, examining phenomena ranging from the growth of underlying phonological knowledge to phonotactic probability effects, repetition accuracy for known and novel words, stimulus length effects, serial position effects, and age effects. Furthermore, the manipulation of PSTM functionality in Simulations 5–6 constitutes, in effect, a simulation of individual differences variation. It is worth highlighting once again that all of these phenomena were simulated with a *single* set of parameters across all simulations (except where the very purpose of the simulation was to explore the effect of variation in a parameter). That is, we did not fit the model differently to different phenomena.

With respect to the theoretical debate that motivated this work, our computational formulation emphasizes three points. First, as posited by the PSTM account, PSTM functionality plays a causal role in the development of phonological vocabulary/knowledge – i.e., in phonological vocabulary learning. It is by virtue of taking seriously the fact that word forms are psychological entities that extend over time and that are serially ordered, and by considering the computational requirements of producing and learning such stimuli, that the necessity for context information and thus PSTM becomes evident. In perspectives in which the role of PSTM in word form production/learning has been de-emphasized, the temporal nature of word forms may have been regarded as an incidental rather than essential characteristic, and computational considerations have perhaps not been highlighted, so that the necessity for PSTM may have appeared less evident. The present work emphasizes the importance of taking into account the serially ordered nature of word forms, and emphasizes that it is this characteristic that leads to the causal role of PSTM functionality in phonological vocabulary learning. Second, as posited by the linguistic account, phonological vocabulary/knowledge is a causal determinant of nonword repetition accuracy. The model exhibits phonotactic probability effects, suggesting that the phonological knowledge it develops is plausible; it also exhibits effects considered to be markers of PSTM involvement in nonword repetition, suggesting that its account of PSTM functionality is plausible. All of this leads to the third major point: the theoretical opposition between the PSTM and linguistic accounts is illusory, and should be dispensed with; the causal links postulated by each account are consistent with each other. As we noted in Introduction, this point of view is certainly not new, being consistent with our own previous work, which has emphasized both of these directions of causality (e.g., Gupta, 1995; Gupta & MacWhinney, 1997; Gupta et al., 2005), as well as with formulations of the PSTM account in which the possibility of both directions of causality is acknowledged (e.g., Gathercole, 2006; Gathercole, Frankish, et al., 1999). The present work, however, goes beyond such informal statements of this view by providing a computational demonstration of *how* the two theoretical positions can be reconciled. We believe that such a computational demonstration makes an important contribution to our understanding of the issues – in much the same way, for instance, that computational models of performance in the immediate serial list recall task are considered to have furthered our understanding of phenomena described by the less formally specified working memory model of Baddeley, Hitch, and colleagues.

Limitations

The present work has numerous limitations, of course. For instance, the model does not incorporate effects of variables such as semantics, lexical stress, low-level perceptual processing, and low-level articulatory processing, all of which are likely to account for some of the variance in human behavior. The lack of semantics in particular is an important limitation that restricts the model to being a model of phonological word form learning, rather than of word learning more broadly. The lack of semantics also prevents the model from providing a full account of phenomena involving dissociations between known word and nonword processing, such as the contrast, in the neuropsychological syndrome termed the "pure STM" deficit, between the relatively spared ability of patients to repeat known words and their severely impaired ability to repeat nonwords (Baddeley, 1993; Baddeley et al., 1988; Vallar & Baddeley, 1984). Also, as noted earlier, the model incorporates within-word sequencing only at the level of syllables, and not at the phoneme level. A further limitation is that word forms processed by the model are limited to four syllables in length. We adopted this limit primarily in order to reduce computing time over the 8000 epochs of training. Although four syllables is a longer word form length than in most other computational models that we are aware of, it is nevertheless a limitation. In addition, the model provides largely qualitative rather than quantitative fits. Here, it is worth noting that the model's quantitative fits could very likely have been improved by adjusting parameters for simulation of different effects, as is quite commonly done in computational modeling work. The qualitative rather than quantitative fits provided by the model are therefore to some extent a cost of our maintaining a single set of parameter values across simulation of several different effects.

As noted above, these various limitations certainly prevent the model from constituting a full account of vocabulary learning (nor, to our knowledge, does any other extant computational model constitute such an account). However, they do not prevent the model from achieving its primary aim, which was to computationally specify the role of PSTM and vocabulary size/knowledge in phonological vocabulary learning and nonword repetition, as a means of examining the extent to which the PSTM and linguistic accounts are necessarily in opposition to each other. It is worth situating this achievement with respect to other computational models that have addressed related issues.

Relation to other models

One recent computational model has sought very directly to examine the relative roles of working memory and vocabulary knowledge in nonword repetition and nonword learning, i.e., phonological vocabulary learning (Jones et al., 2008). This interesting work employs the computational framework of EPAM, a discrimination net (Simon & Feigenbaum, 1964). However, the model's simulation of nonword repetition does not incorporate certain features that, in our view, are critical characteristics. For instance, the model's simulation of nonword repetition does not produce serially ordered output, nor does it produce the nonword *after* input presentation is complete. As we have emphasized in the present work, we take these to be key characteristics of nonword repetition, ones that go to the heart of what makes the task a working memory task. Accordingly, the Jones et al. (2008) model characterizes working memory in a very different way than in the present model. In addition, that model does not appear to actually produce output at all, and the behavioral performance it simulates appears more akin to nonword recognition than nonword repetition/production. For these reasons, we regard the present work as considerably different from that of Jones et al. (2008), and as making a substantial additional contribution.

Another recent model (Gupta, 2009) focuses on the role of PSTM functionality in the commonly observed relationship between nonword repetition and immediate serial list recall, rather than on the role of PSTM functionality in the relationship between nonword repetition and vocabulary-related variables as in the present work. That work presents a rather different operationalization of PSTM functionality in which the key mechanism is one that is external to the language system, although it operates in close cooperation with it. Briefly, this mechanism is a device (a "sequence memory") that produces a time-varying context signal, which is associated to elements that are active at levels of representation in the language system. This serial ordering mechanism is a variant of the *avalanche* (Grossberg, 1978), and time-varying context signals of this general kind have been employed in a number of computational models of verbal short-term memory (e.g., Brown et al., 2000; Burgess & Hitch, 1992, 1999; Gupta, 1996; Hartley & Houghton, 1996; Houghton, 1990; Vousden et al., 2000). When a set of word form units is activated in sequence at the lexical level of linguistic representation in the Gupta (2009) model (as a result of simulated presentation of a word list), the sequence memory encodes that sequence of activations, and can subsequently replay the sequence, thus simulating list recall. Similarly, when a set of sublexical units (such as phonemes or syllables) is activated in sequence as a result of presentation of a word form to the model (as a sequence of such constituents), the sequence memory encodes the sequence, and can subsequently replay it, thus simulating word form repetition. The common involvement of the sequence memory at both these levels of representation provides an account of how PSTM functionality is necessary for both nonword repetition and immediate serial list recall. The Gupta (2009) model also simulates the long-term learning of word forms. However, the sequence memory is not directly involved in this learning, and this is what constitutes the limitation of the model with respect to the issues of present interest: it does not offer an account of the role of PSTM functionality in phonological vocabulary learning. Thus although the Gupta (2009) model succeeds in relating PSTM functionality to nonword repetition and list recall, it does not succeed in relating it to phonological vocabulary learning, and the present work thus makes a substantially different contribution.

Three other models have also addressed issues relating to STM and vocabulary learning, although without addressing the relation between vocabulary size/knowledge and nonword repetition. One of these, developed by Hartley and Houghton (1996), provided an impressive account of nonword repetition, treating a nonword as being comprised of a sequence of one or more syllables, and providing for the encoding and recall of the serial order of phonemes within these syllables. In essence, each syllable incorporated a sequence memory of the general type described above. Repetition of a nonword consisted in activating each encoded syllable successively, in correct serial order, so that the sequence memory contained in each syllable could then in turn spell out its own encoding of the phonemes comprising that syllable, in serial order. Thus the model incorporated the ability to repeat nonwords as sequences of phonemes. It also provided a detailed account of errors in repetition of lists of monosyllabic nonwords. It operationalized STM as the sequence memory contained within each syllable (there was also assumed to be a sequence memory and hence STM that operated to retain the order of a sequence of syllables in the case of a polysyllabic word form), and showed clearly why STM is necessary for nonword repetition. This model was the first to tackle empirical data regarding nonword repetition, and has been quite influential. However, Hartley and Houghton (1996) did not simulate how a phonological vocabulary could be learned, and consequently, did not offer a computational account of the relationship between vocabulary-related variables (such as phonotactic probability) and PSTM functionality or nonword repetition, and the model therefore does not speak directly to the debate between the linguistic and PSTM accounts. An earlier model (Houghton, 1990) did simulate the repetition of novel word forms, and their long-term learning as chunks that incorporated the serial order of their constituents. Each of the chunks in effect incorporated a sequence memory, and the model thus offered an account of the role of STM in nonword repetition and word learning. However, because the chunks were separate, the phonological knowledge stored in the connection weights for one chunk did not interact with that in other chunks, and this knowledge therefore did not influence the repetition or learning of additional word forms. The Houghton (1990) model thus also did not speak to the theoretical debate of focus in the present work. The same points apply to computational work by Gupta (1995, 1996), which took as its starting point the Hartley and Houghton (1996) and Houghton (1990) models.

Clarification of constructs

The present model provides a clear operationalization of a number of constructs that have been central in the literature in discussion of the issues at hand. One such key construct is "phonological short-term memory" or PSTM, which we have preferred to term *PSTM functionality*. As we have emphasized, (P)STM functionality is crucial to the production of serially ordered behavior, of which word form production is an example. As also discussed, and as emphasized by Botvinick and Plaut (2006), the functionality of (P)STM depends (in the present model, as in this type of recurrent neural network in general), on the recurrent connections on the hidden layer. The transmission of information from previous time steps to the current time step occurs via these recurrent connections; the provision of such information is precisely the functionality of short-term memory. The present work makes explicit the nature of this functionality, highlighting the fact that it is influenced by long-term knowledge, because the strength of the weights on the recurrent connections is determined via the long-term learning weight adjustment procedure. At the same time, the constructs of PSTM functionality and long-term knowledge are *not* the same, and our model provides an operationalization of this as well, showing both how they are related and how they are different. We have also attempted to emphasize the distinction between the construct of PSTM functionality, which we view as the appropriate conceptualization, and the construct of PSTM as an isolable entity or structure or process, which we suggest is a less productive way of thinking about this cognitive ability.

Relatedly, PSTM *capacity limitations* in the model do not arise from any one isolable entity or sub-component. Rather, such capacity limitations (as measured, for instance, by the model's repetition accuracy for longer nonwords) emerge from the functioning of everything in the model that affects the ability to provide information about the past. Variation in the efficiency of activation transmission between units, for instance, would affect PSTM functionality. So would variation in the learning rate (because it would affect connection weights, on the recurrent connections and elsewhere). So would variation in structural aspects of the model such as the number of hidden units. As the overall functioning of the model is thus *multiply determined* (a notion that has recently been emphasized, e.g., Gathercole, 2006; Gupta, 2006, 2008), so is its PSTM capacity.

The present work also provides an operationalization of the relationship between what is frequently referred to as *working memory* (meaning information maintenance) and *statistical learning* (the type of long-term learning incorporated in the model, which can also be viewed as *procedural learning*; see Gupta & Cohen, 2002). As we have emphasized, PSTM functionality, being intrinsically colored by long-term knowledge, is affected by anything that affects long-term knowledge – such as procedural learning – but is nevertheless a distinct functionality. In terms of Fig. 2, procedural learning functionality is what provides for incremental connection weight adjustment and hence learning throughout the system. PSTM functionality is what provides for availability of information to guide serially ordered processing of word forms in the system. As the many structures and processes that provide for PSTM functionality include ones that change via procedural learning functionality (especially the connections from the context layer to the hidden layer), PSTM functionality is affected by procedural learning functionality – but is not subsumed by it.

Four further constructs that are given a clear operationalization in the present work are those of phonological vocabulary *knowledge*, phonological *vocabulary*, phonological vocabulary *size*, and phonological vocabulary *learning*. Phonological vocabulary knowledge is defined by the current state of connection weights in the system. Phonological vocabulary is defined as the set of phonological words in the language that the model can produce correctly in repetition (which will depend on the phonological knowledge, but also on processing factors in the model). Phonological vocabulary size is the *number* of items in the vocabulary set. Phonological vocabulary learning is defined as an increase in vocabulary size between two points in time. In the model, these are all related but nevertheless separable constructs. By clarifying this, the model encourages treatment of them as separate variables in theoretical discussion. These distinctions also apply, of course, to vocabulary in general (i.e., not just to phonological vocabulary, but also to vocabulary knowledge that includes semantic content).

Conclusion

As noted in Introduction, the causal involvement of PSTM functionality in development of phonological vocabulary/knowledge has potentially profound implications. By providing a computationally specified account of this involvement, therefore, the present work offers insight into the foundational human ability to learn words, indicating that this element of linguistic ability, which is often considered uniquely human, is founded on a *memory* mechanism (or functionality). It also suggests that in the phonological domain, short-term and long-term memory functionalities are inextricably linked in their effect, but still separable in their mechanisms. Given the computational considerations underlying this last point, it would be somewhat surprising if this principle did not apply to non-phonological domains as well, such as visual short-term memory, and this extends the potential implications of the present work considerably.

We therefore believe that the present model takes a worthwhile step toward understanding aspects of phonological vocabulary learning and resolving a theoretical debate of two decades, while also offering ideas about the relationship between short-term and long-term memory functionality in other cognitive domains.

Acknowledgments

We wish to thank Matt Botvinick and Dave Plaut for generous assistance in using the Botvinick and Plaut (2006) model as a starting point for the present work, as well as helpful discussion, Doug Rohde for many helpful clarifications regarding use of the Lens simulator, and Gary Dell, Brandon Abbs, Libo Zhao, Stephanie Packard, and Myrna Schwartz for helpful comments on a draft of this article. This research was supported in part by Grants NIH NIDCD R01 DC006499 to Prahlad Gupta, and NIH NIDCD R01 DC01924 to Nadine Martin, Prahlad Gupta, and others.

References

Archibald, L., & Gathercole, S. (2007). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics, 28*, 587–606.

Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19*, 537–552.

Baddeley, A. D. (1993). Short-term phonological memory and long-term learning: A single case study. *European Journal of Cognitive Psychology, 5*, 129–148.

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., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586–595.

Black, A. W., & Taylor, P. A. (1997). The festival speech synthesis system: System documentation (Technical Report HCRC/TR-83). Scotland, UK: Human Communication Research Centre, University of Edinburgh. <http://www.cstr.ed.ac.uk/projects/festival.html>.

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

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year olds. *Journal of Experimental Child Psychology, 63*, 44–78.

Brown, G. D. A., & Hulme, C. (1996). Nonword repetition, STM, and word age-of-acquisition. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 129–148). Hove: Psychology Press.

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

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*, 429–460.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551–581.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*, 627–652.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234–272.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*, 221–268.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17*, 149–195.

Dollaghan, C. (1987). Fast mapping in normal and language-impaired children. *Journal of Speech and Hearing Disorders, 52*, 218–222.

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

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.

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology, 46A*, 533–558.

Ellis, N., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology, 49A*, 234–250.

Ellis Weismer, S., & Edwards, J. (2006). The role of phonological storage deficits in specific language impairment: A reconsideration. *Applied Psycholinguistics, 27*, 556–562.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Ernansky, M., Makula, M., & Benuskova, U. (2007). Organization of the state space of a simple recurrent network before and after training on recursive linguistic structures. *Neural Network, 20*, 236–244.

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

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., & Baddeley, A. D. (1990a). Phonological memory deficits in language-disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336–360.

Gathercole, S. E., & Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: A study of young children learning arbitrary names of toys. *British Journal of Psychology, 81*, 439–454.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84–95.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Short-term memory and new word learning in children. *Developmental Psychology, 33*, 966–979.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*, 65–77.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12*, 349–367.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887–898.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.). *Progress in theoretical biology* (Vol. 5). New York: Academic Press.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.). *Pattern recognition by humans and machines: Speech perception* (Vol. 1). New York: Academic Press.

Gupta, P. (1995). Word learning and immediate serial recall: Toward an integrated account. Ph.D. thesis, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA.

Gupta, P. (1996). Word learning and verbal short-term memory: A computational account. In G. W. Cottrell (Ed.), *Proceedings of the eighteenth annual meeting of the cognitive science society* (pp. 189–194). Mahwah, NJ: Lawrence Erlbaum.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology A, 56*, 1213–1236.

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13*, 318–324.

Gupta, P. (2006). Nonword repetition, phonological storage, and multiple determination. *Applied Psycholinguistics, 27*, 564–568.

Gupta, P. (2008). The role of computational models in investigating typical and pathological behaviors. *Seminars in Speech and Language, 29*, 211–225.

Gupta, P. (2009). A computational model of nonword repetition, immediate serial recall, and nonword learning. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 108–135). Hove, England: Psychology Press.

Gupta, P., & Cohen, N. J. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological Review, 109*, 401–448.

Gupta, P., Abbs, B., Tomblin, J. B., & Lipinski, J. (in preparation). Phonological priming, phonological perception, phonological STM, nonword repetition, and word learning in SLI.

Gupta, P., Lipinski, J., Abbs, B., & Lin, P.-H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language, 53*, 141–162.

Gupta, P., Lipinski, J., Abbs, B., Lin, P.-H., Aktunc, M. E., Ludden, D., et al. (2004). Space aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behavioral Research Methods, Instruments, and Computers, 36*, 599–603.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*, 267–333.

Gupta, P., MacWhinney, B., Feldman, H., & Sacco, K. (2003). Phonological memory and vocabulary learning in children with focal lesions. *Brain and Language, 87*, 241–252.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*, 1–31.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation*. New York: Academic Press.

Jones, G., Gobet, F., & Pine, J. (2008). Computer simulations of developmental change: The contributions of working memory capacity and long-term knowledge. *Cognitive Science, 32*, 1148–1176.

Jordan, M. I. (1986). Serial order: A parallel distributed processing approach (Report 8604). San Diego, La Jolla, CA: Institute for Cognitive Science, University of California.

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes*. Mahwah, NJ: Lawrence Erlbaum.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics, 21*, 3–28.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. RI: Brown University Providence.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1–36.

Martin, N., & Saffran, E. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*, 641–682.

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language, 52*, 83–113.

Massaro, D. W. (1989). *Experimental psychology*. Orlando, FL: Harcourt Brace Jovanovich.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375–405.

Menyuk, P., Menn, L., & Silber, R. (1986). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (2nd ed.). Cambridge, England: Cambridge University Press.

Metsala, J. L. (1999). The development of phonemic awareness in reading disabled children. *Applied Psycholinguistics, 20*, 149–158.

Michas, I. C., & Henry, L. A. (1994). The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology, 12*, 147–164.

Montgomery, J. W. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology, 11*, 77–91.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research, 44*, 778–792.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetition 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.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language learning. *Journal of Memory and Language, 30*, 331–347.

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effects of phonological similarity and item length. *Quarterly Journal of Experimental Psychology, 44A*, 47–67.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (Eds.). *Parallel distributed processing: Foundations* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Saffran, E. M. (1990). Short-term memory impairment and language processing. In A. Caramazza (Ed.), *Cognitive neuropsychology and neurolinguistics*. Mahwah, NJ: Lawrence Erlbaum.

Santos, F. H., Bueno, O. F. A., & Gathercole, S. E. (2006). Errors in nonword repetition: Bridging short- and long-term memory. *Brazilian Journal of Medical and Biological Research, 39*, 371–385.

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

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45A*, 21–50.

Service, E., & Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language, 32*, 608–623.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155–172.

Simon, H. A., & Feigenbaum, E. A. (1964). An information-processing theory of some effects of similarity, familiarization, and meaningfulness in verbal learning. *Journal of Verbal Learning and Verbal Behavior, 3*, 385–396.

Snowling, M. J. (2006). Nonword repetition and language learning disorders: A developmental contingency framework. *Applied Psycholinguistics, 27*, 588–591.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on gathercole, willis, emslie, and baddeley. *Applied Psycholinguistics, 12*, 369–373.

Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior, 23*, 151–161.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavioral Research Methods, Instruments, and Computers, 36*, 481–487.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*, 101–175.

Wagner, K. R. (1985). How much do children say in a day? *Journal of Child Language, 12*, 475–487.