Learning on the fly
Computational analyses of an unsupervised online-learning effect in artificial grammar learning

Martin Rohrmeier (mrohrmeier@cantab.net)
Centre for Music & Science, Faculty of Music, University of Cambridge

Abstract

Humans rapidly learn complex structures in many domains. Some findings of above-chance performance of untrained control groups in artificial grammar learning studies raise the question to which extent learning can occur in an untrained, unsupervised testing situation with partially correct and incorrect structures. Computational modelling simulations explore whether an unsupervised online learning effect is theoretically plausible in artificial grammar learning. Symbolic n-gram models and simple recurrent network models were evaluated using a large free parameter space and applying a novel evaluation framework, which models the human experimental situation through alternating evaluation (in terms of forced binary grammaticality judgments) and subsequent learning of the same stimulus. Results indicate a strong online learning effect for n-gram models and a weaker effect for simple recurrent network models. Model performance improves slightly once the window of accessible past responses for the grammaticality decision process is limited. Results suggest that online learning is possible when ungrammatical structures share grammatical chunks to a large extent. Associative chunk strength for grammatical and ungrammatical sequences is found to predict both, chance and above-chance performance for human and computational data.

Keywords: Unsupervised learning; online learning; computational modelling; artificial grammar learning; n-gram model; neural network; artificial grammar learning

Introduction

Humans are very efficient learners. In many cases we learn without intention and without awareness, and it has been suggested that implicit learning constitutes one powerful and fundamental root mechanism of learning (Reber, 1993). Humans are even further able to learn and to adapt to the environment, *whilst being in the midst of things*: we pick up individual characteristics, or melodic features in a piece of music while we are listening or dancing to it, sportsmen are able to adapt to characteristics of their opponents or the environment while playing, or musicians adapt to characteristic musical patterns of other musicians while improvising together.

Humans acquire implicit knowledge about regular structures very quickly. Serial reaction time experiments have found humans to be able to acquire rule-based structures extremely rapidly (Reber, 1993). Similarly, under the artificial grammar learning paradigm (AGL) participants acquire rule-based structures rapidly after short familiarisation periods (Pothos, 2007). One question that arises in this context concerns how efficient humans may learn regular structures even during a test, or under more complex conditions involving a combination of both, regular and irregular structures. For instance, Dulany et al. (1984) found that untrained controls performed above chance, which might suggest that they have picked up some regularity in the structures during the testing. Redington & Chater (1996) discuss the possibility of such a learning process, whereas Reber & Perruchet (2003) argue that above chance performance of a control group would not stem from a learning effect but from confounding structural biases that may be easy to detect. However, two recent musical grammar learning experiments found a high performance of about 60% in untrained controls (Loui et al, 2008; Rohrmeier et al., submitted) which may reopen the question about a potential rapid online-learning effect.

This study addresses how online-learning on the fly could be theoretically possible based on computational modelling methods. It proposes a framework to model both the simultaneous learning of structures while being tested and the generation of binary grammaticality judgments, in a way that parallels the human situation. It aims to demonstrate that two standard computer models of learning reproduce an effect of unsupervised online learning under certain conditions regarding the stimulus structures. Further it explores why it turns out that grammatical structures, but not ungrammatical ones, are preferred as familiar even though the learning process happens under unsupervised conditions. These theoretical and computational observations raise several hypotheses regarding an efficient online-learning effect for future psychological research.

Experimental hints & evidence

In a musical AGL experiment, Rohrmeier et al. (submitted) found that untrained control participants were able to distinguish rule-consistent grammatical stimulus structures from ungrammatical structures throughout the course of a testing phase, even though they had no prior training. Once the performance of this group is plotted over time (throughout the course of the testing phase, in which the stimulus order was randomized), one finds a curve of the shape of a saturation curve (figure 1). The fact that the performance curve begins at a chance level of 0.5 (and not above) and steadily raises to a level of 0.62, suggests that participants gradually pick up some knowledge that enables them to distinguish the structures, with little prior bias. The study found the group performance to be significantly above chance after 11 steps into the testing phase.

This unusual result is surprising and rare in the context of other AGL studies. However there are not many cases of studies with untrained control groups. Dulany et al. (1984), Redington & Chater (1994), Dienes (reported in Redington

& Chater, 1996), and Loui et al. (2008) found above chance performance of untrained controls; whereas Altmann et al. (1995), Meulemans & Van der Linden (1997) and Reber & Perruchet (2003) did not. If well this set of experimental evidence is not decisive and further empirical work is required, computational modelling work may shed light on the question of whether an effect of rapid online-learning under complex conditions of partially grammatical and ungrammatical structures is theoretically plausible at all. In addition it may raise particular hypotheses regarding human learning performance based on theoretical considerations.

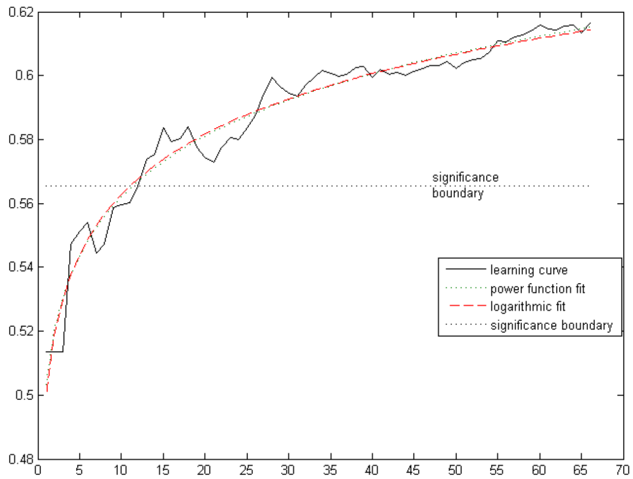

Figure 1: Performance of an untrained participant group during the testing phase.

Based on these considerations, this study aims to simulate a potential effect of online-learning from the angle of two different cognitively motivated models: a connectionist model with reference to connectionist theories of AGL (Pothos, 2007) and a symbolic n-gram model with reference to fragment or chunking based theories of human learning (Servan-Schreiber & Anderson, 1990; Perruchet & Pacteau, 1990).

Method

First, the modelling framework intends to model the simultaneity of learning and responding during testing. This departs from traditional machine learning or computational modelling methods (Mitchell, 1997; Bishop, 2006) as the typical separation between model training and model evaluation is suspended. In this framework the models are first evaluated for each given stimulus and then subsequently trained on the same stimulus. This method keeps the modular operations of training and evaluating the model with single strings (as learning during the processing of the stimulus would require significant changes in the mechanism of the model, in particular, the SRN).

Secondly, the modelling framework intends to capture the human testing situation, which involves having to decide about stimulus grammaticality immediately during the testing. Often computational models are simply evaluated by comparing the overall sequence familiarity for grammatical and ungrammatical sequences after the whole test evaluation (e.g. Kuhn & Dienes, 2008) but are not required like the human to give decisive binary grammaticality (G/UG) responses after each single stimulus without full information about the remaining test set. Consequently, the model responses would not be directly comparable to the human responses. Therefore, the present modelling framework applies a threshold decision technique to generate binary grammaticality judgments from the model's familiarity responses directly for each single stimulus (see below).

We use cross-entropy based on sequence predictability (Mitchell, 1997; Bishop, 2006; Pearce & Wiggins, 2004) as an estimate of the familiarity that a model assigns to a stimulus.

Models

N-gram model. Fragment based n-gram models are symbolic models which have been successfully used in computational linguistics and in music modelling (Manning & Schuetze, 1999; Pearce & Wiggins, 2004, 2006). This study employs a simple n-gram model after Pearce & Wiggins (2004) which stores fragments of the lengths 1 to n symbols from its input sequences, and creates predictions for the symbol sequence of a given test sequence by combining predictions from differently sized fragments using Moffat's (1990) method, which has been found to perform best in comparison to other smoothing and combining methods (Pearce & Wiggins, 2004). The model produces a familiarity response for a whole test sequence based on its information content, i.e. the mean cross-entropy of the prediction for each symbol of the sequence.

Simple Recurrent Network. The simple recurrent network model was implemented following Elman (1990). A familiarity response for a single test sequence is generated through the information content, i.e. cross-entropy based on the prediction of each symbol.

Deciding grammaticality judgements

Both models return familiarity values based on cross-entropy, which have to be classified on the fly into binary grammaticality responses. As the range and distribution of the familiarity values are unknown prior to the test and vary over time, the decision cannot be based on a static threshold value. The current familiarity value is instead classified as grammatical or ungrammatical when it is greater or smaller than the median of the available past familiarity values. The decision is made random for the first sequence as there is no reference value available.

Procedure

First the model is initialised and the sequence order is randomised. Then, for each stimulus of the testing set, the model computes, as outlined, a familiarity response based on cross entropy, which is compared to the median of the past responses and subsequently transformed into a grammaticality judgment. After each sequence evaluation, the model is trained with the stimulus.

Choice of free parameter space

Cleeremans & Dienes (2008) discuss the problem that regarding the choice of free model parameters there are few ways of determining cognitively meaningful parameter choices. The present simulations adopt the method by Kuhn & Dienes (2008) to define a grid over the range of possible meaningful parameters and to run a fixed number of simulations for each point in the parameter space. A parameter space of learning rate and momentum each of {0.1, 0.3, 0.5, 0.7, 0.9}, 2 learning epochs, and {10, 15, 25, 50, 80, 120} hidden units was used for the SRN models, resulting in a space of 150 parameter combinations. The n-gram models were evaluated using a parameter space of a maximal n-gram length of $\{2,3,4,5,6,\infty\}$, where ∞ signifies that there was no upper limit for the fragment size and that fragments up to the whole string were stored.

Materials

Test sequences from the studies above which featured an untrained control group were used, if the stimuli were available. In addition, the stimuli by Brooks & Vokey (1991) as used by Tunney & Shanks (2003) were included in order to feature another well-known finite-state grammar.

Simulation 1

The purpose of simulation 1 was to investigate to which extent online learning could be simulated for the studies listed above. For each of the 7 grammars listed above, 80 instances of each the n-gram model and the SRN were run for each configuration in the parameter space above. In addition, the same number of control models were run, which featured no sequence training after stimulus presentations.

Table 1 displays the results. All n-gram models exhibit a significant and strong effect of online-learning for all parameters (all $p<0.0005$). In many cases mere bigram learning proves sufficient for a performance level which is barely topped by larger contexts, a finding that is consistent with evaluations by Pearce & Wiggins (2004). Further, many n-gram models outperform human results. SRN models also show significant above chance performance, typically for 50 or more hidden units and a learning rate of 0.5 or higher. All control models performed not different from chance (all $df=79$, $p>0.05$) for all stimulus sets, suggesting that there was no model induced bias. In general, the SRN models tend to have a less strong effect of online-learning and often perform slightly lower than humans. However, unlike many n-gram models, SRN models exhibit around chance performance for the stimulus set by Reber & Perruchet (2003), just like in the human results. The structures by Meulemans & Van der Linden, exp. 2a were not learned by either models or humans, whereas in their exp 2b, interestingly, models and humans preferred ungrammatical structures as familiar.

Simulation 2

The purpose of simulation 2 was to investigate to which extent the window of available past familiarity judgments influences the online-learning efficiency. Therefore, one small change was introduced to the process of the grammaticality judgement decision: whereas the grammaticality response compared the current familiarity value to all previous familiarity values, now it was only compared to the last 5, 10, 20, or 30 values, using a sliding window technique. The cognitive motivation for this change was to incorporate some of the effect of human memory limitations in the modelling.

The same models and the same parameter space as in simulation 1 have been evaluated for the different memory windows above. Results revealed that performance for both model types slightly improved overall when less (window size of 10 or 20) but not too little context (window size of 5) of familiarity judgments is taken into account. The mean model performance improved for .003, .010, .013, .007 (n-gram models), and 0.012, 0.016, 0.014, 0.009 (SRN models) percent points for memory windows of 5, 10, 20, 30 respectively, compared to an unlimited memory window[1]. This small improvement may be explained through the fact that familiarity values tend to increase and to converge throughout the test. When the familiarity window excluded older values in which the models were in a prior, less stable state, the performance improves, having an even greater effect for high-performing models[1].

Why do the right structures get picked?

The behavioural and computational findings beg the main question of how it is possible that grammatical structures may potentially be learned gradually and in an unsupervised manner, within an environment that contains 50% ungrammatical structures, i.e. a fair amount of misleading and wrong information. The model simulations give rise to a potential explanation and a hypothesis for human behaviour extending Redington & Chater's (1996) argument: stimulus structures, both grammatical and ungrammatical structures, share a large set of fragments or chunks, and those are acquired with every testing of grammatical and ungrammatical stimulus. If one assumes that the learning of chunks or fragments constitutes one major part in artificial grammar learning (Servan-Schreiber & Anderson, 1990; Perruchet & Pacteau, 1990; Pothos, 2007), the chunk distribution of stimuli would supposedly play a major role in the learning. Whereas grammatical chunks appear relatively frequently, ungrammatical chunks, however, arise from violations in the structure and are thus expected to appear less frequently. Once a learner detects differences between chunk frequency in stimuli, a distinction between grammatical and ungrammatical chunks might be possible on that base. Therefore, the reason why responses converge toward grammatical structures may rely on the fact that grammatical sequences tend to have higher chunk

[1] Detailed results had to be omitted out due to space limitations.

frequencies on average than ungrammatical sequences.

Accordingly, one might hypothesise that if grammatical and ungrammatical chunks were to appear comparably frequently in the whole test set, the learner could not distinguish between them. Secondly, it would be expected that the learner picks the structures with the larger share of frequent fragments as grammatical; and hence the selection would converge toward either grammatical or ungrammatical structures depending on which one encompasses the more frequent chunks. Using the associative chunk strength (ACS) measure (Meulemans & Van der Linden, 1997), we would predict that the set of stimuli with the greater mean ACS with respect to the whole set of testing structures will be preferred and that the performance would be around chance if both mean ACS values were very similar.

Accordingly, the proportion of mean grammatical ACS to ungrammatical ACS was calculated for the different stimulus sets used above. The ACS proportion values were roughly about 1 for Meulemans & van der Linden, exp 2A, Reber & Perruchet; greater than 1 for Dulany et al., Loui et al., Rohrmeier et al., and Tunney & Shanks, and smaller than 1 for Meulemans & van der Linden, exp 2B. Mean ACS values for grammatical and ungrammatical structures were significantly different for Dulany et al, Rohrmeier et al., Tunney & Shanks Meulemans & van der Linden, exp 2B (all $p<0.02$). Both human performance and model performance match the pattern of the ACS proportions in terms of both direction and extent of performance: Human performance for the first (balanced) studies is at chance, and models perform not as well or at chance. Human and machine performance for the second set of studies is above chance. In the third case, human performance is below chance (Meulemans & van der Linden (1997) do not report if it is significant) and this is matched by significant below chance performance of the computational models. The correlation between ACS proportions and human as well as model performance were high: 0.71 (human performance), 0.98 (2-gram & 3-gram models), greater than 0.90 (other n-gram models), greater than 0.84 (SRN models with 80 or 120 hidden layers and learning rates greater than 0.7), and 0.89 (best SRN model). Finally, it is interesting to note that n-gram models show that some above chance online learning was possible for the structures by Reber & Perruchet, and Loui et al., even though the mean ACS values for their grammatical und ungrammatical structures were not significantly different (both $p>0.4$).

The learning curve

Another related question concerns the shape of the learning curve. Assuming that the performance curve of the online learning effect mainly depends on the gradual acquisition of information (about the distribution of the stimulus features or chunks) throughout the testing phase, a very simple estimate of the learning and its growth can be formulated based on common considerations. Assuming that new information gained about the sequences decreases as more sequences are known, a decreasing function of information intake may be expressed:

$$f(t) = b \cdot x^{-a} \quad \textit{for} \; \mathrm{a, b} \in \mathrm{R}^{+} \tag{1}$$

Accordingly, the total knowledge about the structures at a certain time step is the amount of the information acquired up to that time:

$$K(t) = \int f(t)\, dt \tag{2}$$

Further, assuming that the performance in term of the likelihood of a correct response is proportional to the total knowledge about the sequences at a time, simple performance curve estimates can be derived:

$$P(t) \propto K(t) \tag{3}$$

$$P(t) = c \cdot K(t) = c\int f(t)dt = \begin{cases} C + k_1 x^m & a \neq 1 \\ C + k_2 \log(t) & a = 1 \end{cases}$$

$$\textit{with} \quad k_1 = \frac{cb}{1-a} \quad k_2 = cb \quad m = 1-a \tag{4}$$

This consideration yields a logarithm or power function prediction, based on two or three free parameters, for the performance curve of the online learning effect. These curves relate to well-known power laws of human learning (Newell & Rosenbloom, 1981; Anderson, 1995) and fit the human data well, which was available for the study by Rohrmeier et al. (Fig. 1). They also match the computational learning curves (Fig. 2) well (all $R^2>0.94$; further details were omitted due to space limitations).

Discussion and Conclusion

The findings above suggest that there are some theoretical and empirical grounds to assume an online learning effect. The results from the first and second simulation show that the online-learning effect can be reproduced by cognitively motivated symbolic and connectionist models and that a limited memory window improves the performance.

The learning effect is possible when ungrammatical structures contain grammatical fragments to a large extent. The considerations and simulations suggest that online learning occurs because responses tend to converge towards sequences with high ACS values, independently of them being grammatical or ungrammatical. This yields a hypothesis for future experimental work: behavioural experiments may reveal whether participants indeed would tend to choose structures with high ACS independently of whether they are rule based or not in an online learning situation. Future work may further assess to what extent ACS of grammatical and ungrammatical sequences predicts the direction and extent of human performance well.

Theories of AGL (Pothos, 2007) propose that there are several theoretically plausible forms of the acquired knowledge, such as chunk knowledge, anchor positions, rule knowledge, or, microrules. This research was based on chunk knowledge and showed that it could predict an online-learning to a certain extent. It remains open which effect the other features or factors may have with regards to the online learning effect.

Although the models in this study show an effect of online learning, the results do not fully account for human results: the fragment-based n-gram models tended to learn 'too efficient' and to outperform the human results whereas the SRN models tended to perform worse than human results. From this perspective, strongly n-gram based accounts of human learning (Perruchet & Pacteau, 1992) would require to incorporate explanations of lower human performance compared to the efficiency of models based on n-gram representations, whereas connectionist accounts would need to account for the better human performance.

One remaining question concerns why this effect has not been commonly found in other studies. The reason why Reber & Perruchet (2003) have found no online learning effect of untrained controls in their experiments, appears to stem from the fact that their grammatical and ungrammatical structures are highly balanced in terms of their ACS. Other studies, in which ACS was unbalanced towards grammatical structures or ungrammatical structures found performance in favour of potential online learning. Yet more experimental evidence is needed.

Another potential explanation for the little present evidence of the effect may be that unambiguously clear control group instructions are difficult to generate and that stimulus appearance might influence learnability in the context of online learning where very quick memorisation is required. Most AGL studies use abstract letter sequences such as VNRX which have little overlap with everyday structures, language, or sounds. In this respect it is striking that two studies which used melodies of simple sequential structure (Rohrmeier et al., submitted; Loui et al., 2008) found very high performance of untrained controls about 60%. Similarly, Reber & Perruchet's (2003) study found higher performance when using consonants common in French language. Whether there is an effect of stimulus domain and appearance for online learning remains to be further explored. These findings have an impact for the AGL research paradigm in as much as some learning effect during testing has to be assumed, even though its additional impact after a learning phase might be small.

Acknowledgments

This work was supported in part by Microsoft Research through the European PhD Scholarship Programme and the Arts & Humanities Research Council.

References

Altmann, G.T.M., Dienes, Z., and Goode, A. (1995). On the modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 899–912.

Anderson, J. R. (1995). *Learning and Memory: An Integrated Approach*. John Wiley & Sons.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Brooks, L.R., and Vokey, J.R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Matthews et al. (1989). *Journal of Experimental Psychology,* 55: 316-323.

Cleeremans, A., & Dienes, Z. (2008). Computational models of implicit learning. In R. Sun (Ed.), *Cambridge Handbook of Computational Psychology*. Cambridge University Press, 396-421.

Dulany, D. E., Carlson, R. A., and Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General,* 113:541-555.

Dienes, Z., Broadbent, D. E., and Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 17, 875-882.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 16*, 41–79.

Kuhn, G., & Dienes, Z. (2008). Learning non-local dependencies. *Cognition*, 106, 184-206.

Loui, P., and Wessel, D. (2008). Learning and linking an artificial musical system: Effects of set size and repeated exposure. *Musicae Scientiae*, 12(2), 207-230.

Manning, C.D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Meulemans, T., and Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23:1007–1028.

Mitchell, T. (1997). *Machine Learning*, McGraw Hill.

Moffat, A. (1990). Implementing the PPM data compression scheme. *IEEE Transactions on Communications*, *38*(11), 1917–1921.

Newell, A. and Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In R, A. J., editor, *Cognitive Skills and their Acquisition*, pages 1–51. Erlbaum, Hillsdale, NJ.

Pearce, M. T., and Wiggins, G. A. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research,* 33(4), 367-385.

Pearce, M. T. and Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, 23(5):377–405.

Perruchet, P. and Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, 119:264–275.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, 133(2):227–244.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* New York: Oxford University Press.

Reber, R., and Perruchet, P. (2003). The use of control groups in artificial grammar learning. *The Quarterly Journal of Experimental Psychology*, 56A: 97–115.

Redington, M., and Chater, N. (1994). The guessing game: A paradigm for artificial grammar learning. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society* (pp. 745-749). Hillsdale, NJ: Erlbaum.

Redington, M., and Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General,* 125:123-138.

Rohrmeier, M., Rebuschat, P., and Cross, I. (submitted). Incidental learning of melodic structure.

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

Tunney, R. T. and Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition*, 31(7):1060–1071.

Evaluation	Parameters		Dulany et al, 1984	Reber & Perruchet, 2003	Loui et al, 2008	Rohrmeier et al, submitted	Meulemans & Van der Linden, 1997 Exp 2a	Exp 2b	Tunney & Shanks, 2003
ACS proportion	bi- and trigrams		1.372	1.009	1.026	1.223	0.975	0.833	1.101
Human results (untrained controls)			0.560	0.445 0.513 0.490	0.60	0.616	0.490	0.450	–
		max n							
n-gram model		2	0.764*	0.538*	0.587*	0.688*	0.486	0.406*	0.579*
mw = ∞		3	0.769*	0.540*	0.592*	0.686*	0.488	0.397*	0.585*
		4	0.758*	0.566*	0.583*	0.726*	0.485	0.411*	0.569*
		5	0.757*	0.573*	0.579*	0.773*	0.502	0.421*	0.575*
		6	0.760*	0.552*	0.597*	0.798*	0.503	0.431*	0.582*
		∞	0.758*	0.574*	0.587*	0.819*	0.491	0.432*	0.576*
n-gram control			0.500	0.500	0.500	0.500	0.500	0.500	0.500
	hid	**lr**							
SRN models	10	0.1	0.491	0.503	0.505	0.506	0.498	0.491	0.503
m = {0.1,0.3,		0.3. 0.5	0.511	0.503	0.507	0.517	0.490	0.490	0.511
0.5,0.7,0.9}		0.7. 0.9	0.518	0.500	0.512	0.530*	0.493	0.476	0.515
mw = ∞	15	0.1	0.499	0.502	0.505	0.511	0.490	0.494	0.505
for all models		0.3. 0.5	0.517	0.499	0.508	0.528*	0.488	0.482	0.519
		0.7. 0.9	0.531*	0.502	0.520	0.529*	0.487	0.471	0.516
	25	0.1	0.500	0.497	0.498	0.515	0.489	0.488	0.511
		0.3. 0.5	0.520	0.500	0.516	0.531*	0.488	0.470	0.519
		0.7. 0.9	0.538*	0.498	0.519	0.536*	0.486	0.463*	0.527*
	50	0.1	0.512	0.494	0.511	0.516	0.493	0.487	0.514
		0.3. 0.5	0.533*	0.501	0.517	0.534*	0.483	0.465*	0.529*
		0.7. 0.9	0.550*	0.500	0.526*	0.542*	0.474	0.449*	0.536*
	80	0.1	0.513	0.502	0.511	0.514	0.488	0.479	0.520
		0.3. 0.5	0.536*	0.497	0.520	0.534*	0.483	0.462*	0.538*
		0.7. 0.9	0.558*	0.497	0.537*	0.544*	0.473	0.446*	0.539*
	120	0.1	0.520	0.501	0.514	0.530	0.478	0.476	0.524
		0.3. 0.5	0.542*	0.495	0.527*	0.538*	0.470	0.452*	0.538*
		0.7. 0.9	0.566*	0.498	0.543*	0.549*	0.474	0.439*	0.538*
SRN control			0.492	0.499	0.501	0.501	0.490	0.496	0.497
Best scoring SRN mw = 20. m=0.1	120	0.7	0.579*	0.493	0.552*	0.577*	0.481	0.444*	0.538*

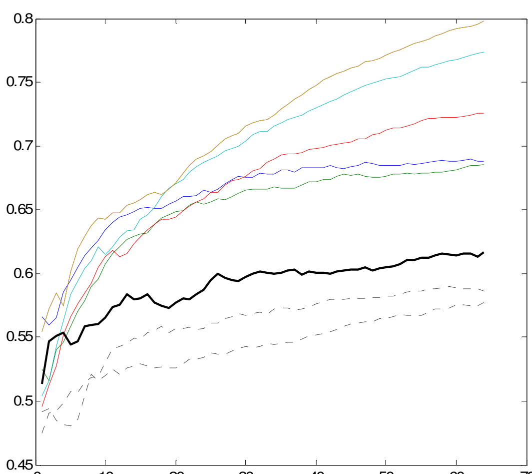

Table 1. Associative chunk strength proportions for bi- and trigrams and mean performance (SD was omitted due to space limitations) for n-gram models and SRN models with no restrictions on the memory window. SRN results were collapsed over all momentum values. All marked (*) mean values are significantly different from chance (all df=79, p<.0001). Displayed parameters are maximal fragment length for n-gram models (max n), number of hidden layer units (hid), learning rate (lr), momentum (m) for SRN models, and memory window size (mw, in number of past stimuli).

Figure 2. Comparing online learning curves for the sequences by Rohrmeier et al. (submitted) for (from top to bottom) n-gram models (coloured) for n=6,5,4,3,2, human performance (thick line) and two high scoring SRN models (dashed, hid=120/80, lr=0.7, mw=20/10, m=0.1/0.7 respectively). Power functions fit all learning curves well (all R^2>0.94), yet plots or details were omitted here due to space limitations.