

Extraordinary Natural Ability: Anagram Solution as an Extension of Normal Reading Ability

Jonathan Henin (jonathan.henin@gmail.com)

Cognitive Science Program, U-1020
Storrs, CT 06269

Emma Accorsi (emma.accorsi@uconn.edu)

E. O. Smith High School
Storrs, CT 06269

Pyeong Whan Cho (pyeong.cho@uconn.edu)

Department of Psychology, U-1020
Storrs, CT 06269

Whitney Tabor (whitney.tabor@uconn.edu)

Department of Psychology and Cognitive Science Program, U-1020
Storrs, CT 06269

Abstract

A recurrent connectionist model of normal word reading, parameterized on an artificial lexicon, can solve anagrams in which the letters of a word have been highly permuted. Distinctive predictions of the model about competition effects in anagram solution and about which reorderings should be hard were supported by two experiments in which human subjects solved English anagrams. The results are in line with prior work on anagram solution, which supports the idea that skilled anagram solvers employ their naturally acquired knowledge of word structure to succeed at this unusual task. The results extend previous modeling efforts by showing how anagram solution ability may be closely related to normal reading ability. Although the model proposed in the current paper is not a learning model, we suggest that exploring the dynamics of recurrent neural networks like the one we propose may provide an avenue by which the theory of cognition can address highly extrapolative generalization ability, in which people excel at tasks that are qualitatively distinct from their training experiences.

Keywords: anagrams; letter scrambling; artificial neural networks; connectionist models; interactive activation; self-organization; expertise; generalization

Introduction

Insight into the nature of human generalization ability is one of the primary achievements of connectionist learning models. In particular, when a network with hidden units has been trained to near zero error on a set of target stimuli, it has often mapped the stimuli into a continuous manifold whose structure reflects the structure of the training environment. Correct predictions about new cases can be made by interpolating on this manifold. But humans sometimes exhibit capabilities that go far beyond their training experience and also, for that matter, plausible conditions of evolutionary shaping. Creative abilities in the arts, athletics, and problem solving may fall into this category. Such cases may be better characterized as “extrapolation” rather than “interpolation”, and indeed, the extrapolation is sometimes of a very radical form. One might ask if connectionist models can provide insight into such cases of radical extrapolation.

The solving of highly permuted anagrams of words by people (e.g., recognizing “rwtae” as a reordering of the word “water”) has in common with such abilities that most people do not solve anagrams extensively while they are growing up; however, in the right conditions, as mature readers, they sometimes experience “pop-out” with an anagram, where the solution to a highly permuted anagram emerges very quickly “without any conscious awareness of a solution attempt” (Novick & Cote, 1992). This may be a case in which a complex behavior develops without explicit training in the behavior. Helpfully, it is an ability that is fairly amenable to psychological study.

Prior Related Work

Prior work on highly permuted anagram solution has revealed several important features of the process:

(a) Pop-out has the hallmarks of parallel constraint satisfaction: it is fast; a person who experiences it cannot typically identify a series of steps involved; it is more common in skilled than unskilled anagram solvers (Novick & Sherman, 2008). (b) Anagrams whose targets are consistent with broad structural patterns in the language are easier to solve than anagrams that are at odds with these, and skilled anagram solvers are more sensitive to the manipulation of these properties than poor anagram solvers (Novick & Sherman, 2008). Relevant structural patterns include bigram frequencies (high bigram frequencies in the target facilitate solution) (Gilhooly & Johnson, 1978; Mayzner & Tresselt, 1958), syllabic patterning (vowel initial words, which are rarer, are harder to solve) (Novick & Sherman, 2008), and phonological regularity (Novick & Sherman, 2008). (c) Anagrams which themselves have low bigram frequencies are easier to solve (Mayzner & Tresselt, 1959) and anagrams whose letter sequence resembles that of the target are easier to solve (Dominowski, 1966; Gilhooly & Johnson, 1978).

An(other) Interactive Activation Model of Visual Word Recognition

These findings would make sense if skilled anagram solution were an extension of natural word-reading ability, which is highly sensitive to structural knowledge of the language and is generally agreed to involve parallel processing in skilled readers. However, we know of no current model of skilled word reading that is able to solve highly permuted anagrams.

In fact, the canonical current models of word recognition, which stem from McClelland & Rumelhart's (1981) Interactive Activation Model of Word and Letter Perception, use slot-filler representations for letters and are thus poorly suited to capture letter permutation effects. We propose a solution that takes up a prominent feature of the early interactive activation models: units corresponding to structures at multiple spatial (and temporal) scales (e.g. phoneme, bigram, word). Unlike in the canonical models, the units in our model detect structure independently of spatial location. They also activate partially in response to partial similarity. As a result, the model is order-sensitive without being order-rigid.

Grimes and Mozer (2001) propose a recurrent connectionist model of anagram solution in which constellations of bigrams compete to be the most active. If the network stabilizes on a pattern which is not a solution, the model identifies a symbol ordering consistent with the current state and uses this symbol ordering to reset the state for another episode of stabilization. This model correctly predicts the positive correlations between high bigram frequency in the target and solution time, and between low bigram frequency in the stimulus and solution time. However, because the constraint satisfaction process is not constrained by lexical information, the model does not seem well-suited to modeling normal reading.

Our model, NGRAMSWELL, modifies the constraint satisfaction architecture of Grimes & Mozer to include a unit for every n-gram in its vocabulary (n ranges from 1 to the length of the longest word). Thus, constraint satisfaction operates over multiple spatial and temporal scales and is simultaneously sensitive to both lexical and sublexical structure. Each unit is connected to every other unit. The units have activations in the range (0, 1), and the vector of activation values, \vec{a} is updated according to

$$\begin{aligned} \frac{d\vec{a}}{dt} &= \text{netinput} \cdot \vec{a} \cdot (1 - \vec{a}) \\ \text{netinput} &= \mathbf{W}\vec{a} + \eta \end{aligned} \quad (1)$$

where \cdot denotes element-wise multiplication, \mathbf{W} is the matrix of weights, and η is small-magnitude Gaussian noise. The weights are given by

$$w_{ij} = \text{Corr}(i, j) \cdot n_i \cdot n_j \quad (2)$$

where $\text{Corr}(i, j)$ is the normalized covariance (correlation coefficient) of n-gram i and n-gram j (two n-grams are counted as co-occurring whenever they appear in the same word). n_x refers to the length, in letters, of n-gram x . The length factors serve to make the interactions between units corresponding to larger n-grams more powerful, thus implementing a

bias toward coherent outcomes. Because the weights are pre-specified, this model does not fit the standard connectionist paradigm for studying generalization, in which a model is trained on some data and tested on others. However, the employment of correlation coefficients as weights is closely related to Hebbian Learning, a well-motivated learning mechanism, and the scaling of weights associated with larger units may be related to the plausible assumption that more neural elements are involved in the conception of more complex objects. We view the current model as a stepping stone to discovering a learning model that addresses the radical extrapolation issue: by identifying the end-state of learning, we may have better luck finding effective architectures and learning mechanisms (Tabor, 2003).

When the model is exposed to a string of letters, each unit in the model activates if its pattern is somewhere in the input (position independent detection) and it activates partially in response to partial match. For example, the input "daisy" activates the $\langle a \rangle$ unit strongly and the $\langle ia \rangle$ unit less strongly. The default initial activation is 0.03; the fully matching units start at 0.04. An n-gram counts as partially matching the input if the input contains its letters in a contiguous group in a different order. In this case, the unit is activated to 0.035 (halfway between the resting level and the full match level). The model settles until the rate of activation change in every unit falls below a threshold (0.0001). For each word, k , in the model's vocabulary, the state in which all the n-grams of word k are maximally activated ($a_i = 1$) and all other n-grams are minimally activated ($a_i = 0$) is called the *ideal state* corresponding to word k . At every point in time, a vector of word activations is computed as the cosine of the angle between the model's current state and the ideal state of each word in its vocabulary. If the target word is not the most activated word when the model reaches the stability threshold, every unit is reset to its initial activation level and the model resettles. The standard deviation of the noise in *netinput* is 0.1. The process repeats until the target is activated or a maximum number of timesteps (800) is reached. We tested the model on small (40-word), artificial vocabularies of 5-letter words generated randomly from an alphabet of 20 letters. Fifteen of the 40 words were constructed to overlap in four out of five letters in order to create a clear contrast between high and low competition words.

If the letters of an actual vocabulary item are presented in the correct order, then the cosine distance vector reliably exhibits an early surge in the activation of the correct candidate, coincident with a plunge in the activations of all other words, and followed by a more gradual stabilization in which the correct candidate has the highest activation and other candidates asymptote at various levels below it. It is this reliable ability to quickly recognize normally-ordered words, in conjunction with our claim made above that the weight structure could be learned under normal circumstances, that supports our claim that this model plausibly approximates normal word reading. On the other hand, if the letters are pre-

sented in a highly permuted order, the activation of the target word sometimes surges, just as it does in normal reading (corresponding, perhaps, to “pop-out” as discussed above). It also often happens that the correct candidate never surges and some other word becomes dominant, or the correct candidate plunges along with the rest of the vocabulary at first while some other word dominates, and then rises slowly to become the preferred choice. If the model stabilizes on an incorrect word before the maximum allowed time has passed, then it is restarted with the same inputs and allowed to stabilize again (corresponding, roughly, to serial solution search). Restarting can remedy a previous failure to discover a solution because the noise in *netinput* causes the model to explore different avenues on each pass. The model has six free parameters: number of letter types, number of words, number of high-overlap words, noise magnitude, stabilization threshold, and maximum time steps. The settings of these parameters (reported above) were chosen via intuitive exploration of the parameter space, guided by the desire to emphasize the contrast in accuracy values between high and low competition words, as defined below.

Predictions

Two predictions of the model distinguish it: (1) Competition/cooperation among similarly structured words should systematically influence accuracy and reaction time; (2) The degree of permutation of the target words should systematically influence accuracy and reaction time.

To make prediction (1) specific and testable, we defined the level- n competition of a word k as

$$JComp_k(n) = 1 - \frac{friends(n)}{sharers(n)} \quad (3)$$

where $sharers(n)$ is the number of words that share some subset of n letters with word k and $friends(n)$ is the number of those that contain n letters contiguously in the same order as they occur in word k . The competition measure allows us to compute a property of real English words that parallels a property of NGRAMSWELL, which we have so far only implemented with small vocabularies. For five letter words of English, most of the variance in competition occurs at level 4 (SDs: Level 2: 0.046; Level 3: 0.053; Level 4: 0.120).¹ Likewise, when NGRAMSWELL is parameterized with a 40 word vocabulary in which 15 words overlapped in four out of five letters, roughly mimicking the high density of overlaps in high-competition English words, most of the variance in competition occurs at level 4 (SDs: Level 2: 0.061; Level 3: 0.097; Level 4: 0.140). Therefore, Experiment 1 (described below) focused on contrasts in $JComp(4)$. $JComp$ is distantly related to neighborhood density measures like *number of orthographic neighbors*, but it posits influences from highly permuted orders, which are not taken into account under the usual 1-letter-difference definition of neighborhood,

¹These statistics were computed from a sample consisting of the words with COBUILD frequency greater than 100 in the CELEX database.

and it differentiates between neighbors that share the letter order of the target word and neighbors that have the same letters in a different order.

To formalize prediction (2), we defined classes of permutations based on Bubblesort, an algorithm which sorts a string by systematically swapping adjacent elements that are out of order (Knuth, 1973).² For five-letter strings with unique letters, the Bubblesort distance from the target (i.e. number of Bubblesort swaps required to transform the string into the target) is 0 for the target itself and 10 for perfect reversal; all other distances lie between these values.

Figures 1 and 2 show the results of a simulation experiment that explored predictions (1) and (2) in NGRAMSWELL. The figures were constructed by choosing a $JComp(4)$ threshold (0.75) that separated the $JComp(4)$ values of the 40 vocabulary items into two distinct clusters termed “High Competition” and “Low Competition”. Each word of the vocabulary was tested 10 times for each of 10 different permutations at each Bubblesort level. The 10 choices at each Bubblesort level were sampled randomly with replacement.³ Two properties of the figures are of central interest here: (i) the mean accuracy of the high competition words is considerably lower than that of the low competition words, while the reading times show the opposite relation, indicating in both cases that anagrams of high competition words are harder to solve; (ii) accuracy and reaction time appear to be in a quadratic relationship with Bubblesort level, indicating that the extreme orders are relatively easy compared to the middle orders.

We ran a multiple regression analysis to test the significance of these claims. Both Bubblesort Level and Competition ($JComp(4)$ value) were analyzed as continuous factors (the High Comp/Low Comp grouping shown in the figures is for illustration purposes only). Indeed, the linear term of $JComp(4)$ was significantly negative ($b = -2.015$, $p < .001$) in the accuracy analysis and significantly positive ($b = 1393.6$, $p < .001$) in the reaction time analysis, supporting the claim that high competition is harder than low competition. Second, the quadratic term of the Bubblesort Level was significantly positive in the accuracy analysis ($b = 0.011$, $p < .001$) and significantly negative in the reaction time analysis ($b = -9.054$, $p < .001$), supporting the claim that the extremes are easier than the middle range. These two findings were probed in experiments with human data, which we discuss below. The locations along the Bubblesort Level axis of the vertices of the fitted parabolas were 6.14 and 5.01 (High vs. Low Competition in the Accuracy data) and 6.27 and 5.16 (High vs. Low Competition in the Reaction Time data). The fact that these values are all greater than 5, the midpoint of the Bubblesort Level range, indicates that the model finds forward order

²Although Bubblesort is notoriously inefficient as a sorting procedure, it provides a simple way of measuring order in anagram stimuli.

³Note that for Bubblesort levels at or near the extreme values, 0 and 10, this method produces many copies of the same structure, but for levels in the middle range, the samples are spread over a variety of orders. When the same order is presented many times, the outcome variation is determined entirely by the noise in *netinput*.

easier than reverse order. We also found a significant interaction between JComp(4) and Bubblesort Level, with a bigger difference between competition levels in the middle range than at the extremes in both accuracy and reaction times ($p < .001$), but we did not probe for this interaction in the human experiments (we leave this as a question for future research). Nine additional simulations with different randomly chosen vocabularies produced the same pattern of results, indicating that the results were not due to random properties of the particular case reported here.

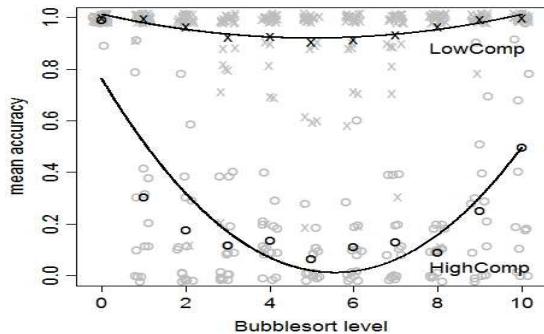


Figure 1: NGRAMSWELL interaction between Bubblesort level and competition level in accuracy. The curves show the best quadratic fit to the individual trial data.

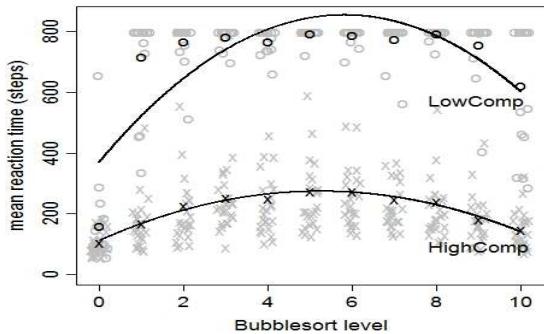


Figure 2: NGRAMSWELL interaction between Bubblesort level and competition level in reaction times. The curves show the best quadratic fit to the individual trial data.

Prediction (1) in NGRAMSWELL stems from the fact that words that share n-grams verbatim with a target tend to boost the activation of those n-grams, thus boosting the target, but words that share permuted n-grams get activated when the target's letters are present and create competition for the target (since the permutations of large n-grams, which have large weights between them, tend to be anti-correlated). Prediction

(2) stems from a general symmetry of linear structures like words: reverse order mirrors forward order. Therefore, each n-gram of the target partially matches an n-gram of the anagram, resulting in a coalition of units that tends to favor the target, though not as strongly as with correct order, because of the reduction of activation with partial match.

Many of the empirical studies on anagram solution to date point to bigram and whole word frequency measures as the strongest predictors of solution accuracy and speed (Gilhooly & Johnson, 1978; Mendelsohn & O'Brien, 1974; Mendelsohn, 1976; Novick & Sherman, 2004, 2008). However, in the absence of explicit modeling proposals, there has been little motivation for considering a measure like JComp(4). We manipulate this measure here in order to see how well it predicts human performance. Two experiments explored predictions (1) and (2), respectively, of the model.

Experiment 1

Method

Participants. 31 college students from the University of Connecticut participated for course credit.

Materials. The set of all words with COBUILD (1991 version) frequency higher than 100 was culled from the Celex database. This set was used to calculate competition values for all the five-letter words in the set at levels 1 through 5. From this set, 10 words were chosen to have $JComp(4) > 0.9$ ($M = 0.94$, $SD = 0.01$) and one anagram solution (HIGH competition, 1 solution), another 10 words were chosen to have $JComp(4) > 0.9$ and two solutions (HIGH competition, 2 solutions), and a third set of 10 had $JComp(4) \leq 0.8$ ($M = 0.63$, $SD = 0.10$) and one solution (LOW competition, 1 solution). All test words had HAL frequency > 1500 (Balota et al., 2007). The HIGH, 1 solution and LOW, 1 solution targets were chosen with a wide range of log frequency values (HIGH: $M = 4.46$, $SD = 0.65$; LOW: $M = 4.32$, $SD = 0.72$), and each HIGH competition word was paired with a LOW competition word of similar frequency. All targets were monomorphemic. Anagram orders were evenly distributed across Bubblesort levels 4, 5, and 6 (the middle of the Bubblesort range). A windows PC was used with a standard keyboard as the input device. The GUI was designed in E-Prime.

Procedure. Participants read instructions which explained anagrams and indicated a 60-second time limit on each problem. Three practice trials with feedback followed. On each trial, an anagram was displayed in the center of the screen in all capital letters. When the participant felt s/he had reached a solution s/he pressed the spacebar, the anagram disappeared, and s/he had the opportunity to type in a solution, pressing spacebar when s/he was ready to move to a new trial. Each participant attempted 30 critical trials. No feedback was provided on critical trials. The computer recorded time to the first spacebar press on each trial as well as the typed solution.

Results

Solution accuracy was assessed by having a computer compare the typed solution to the intended solution or solutions for each problem. Solutions in the HIGH competition, two solutions condition were counted as correct if they matched either target word. The means and standard deviations of the conditions are shown in Table 1. Since the focus of the present study is the contrast between HIGH and LOW competition, 1 solution problems, we will not discuss the 2 solutions condition further.

An ANOVA with Subject and Item as random factors revealed a significant effect of competition (HIGH vs. LOW) in both accuracy ($F_1(1, 30) = 132.78, p < .001$; $F_2(1, 18) = 34.09, p < .001$) and reaction times ($F_1(1, 30) = 69.82, p < .001$; $F_2(1, 18) = 17.48, p < .001$). The same main effect obtained when we re-ran the reaction time analysis on accurate trials only.

Table 1: Experiment 1 means and standard deviations of accuracy (Acc) and solution time (RT) (human subjects experiment).

Condition	Acc	SD	RT (s)	SD
HIGH, 1 solution	0.41	0.49	29.59	22
LOW, 1 solution	0.80	0.40	16.31	18
HIGH, 2 solutions	0.64	0.48	20.04	19

To make a preliminary comparison between JComp(4) and other predictors of anagram solution statistics, we conducted a stepwise multiple regression with HAL Frequency, Position-Dependent Sum of Bigram Frequency ("Summed Bigram Frequency" or SBF), and JComp(4) as linear predictors in that order. SBF values were obtained from the English Lexicon Project (Balota et al., 2007). In the Accuracy data, HAL Frequency alone accounted for 22% of the variance. Adding SBF significantly improved the model, accounting for an additional 12% of the variance. Adding JComp(4) further improved the model, capturing an additional 22% of the variance. In the final model, JComp(4) and HAL Frequency contributed unique predictive power, but SBF did not. In the Reaction Time data, neither HAL Freq nor SBF captured significant variance. JComp(4) made a marginally significant addition to the contributions of these two.

Discussion

The results confirm the prediction of NGRAMSWELL that anagrams with high competition targets should be harder to solve than anagrams with low competition targets. The analysis also provided a suggestion that JComp(4) has predictive power beyond that of some previously studied measures. We consider this comparison preliminary because the stimuli were designed to maximize the contrast in JComp(4) and Hal Frequency but not SBF.

Experiment 2 probed prediction (2) of NGRAMSWELL: that difficulty should be a quadratic function of Bubblesort distance.

Experiment 2

Method

Participants. 73 college students from the University of Connecticut participated for course credit.

Materials and Design. The experiment manipulated one factor, Bubblesort distance, which ranged from 0 to 10. The stimuli were chosen using the same criteria as in Experiment 1, with the additional constraint that no word could have repeated letters. Eleven lists were constructed using a Latin Square design so that every target appeared at every Bubblesort level across the lists.

Procedure. The procedure was the same as in Experiment 1 except that the instructions were modified to alert participants to the possibility that some trials would require no reordering of letters (Bubblesort distance = 0).

Results

We fitted a quadratic polynomial to the accuracy and the reaction time data, plotted as a function of Bubblesort Distance (Figure 3). A regression analysis indicated that the quadratic term was significantly positive in the accuracy analysis ($b_2 = 0.0073, p < .001$) and significantly negative in the reaction time analysis ($b_2 = -199.82, p < .001$). Consistent with the results from NGRAMSWELL the Bubblesort Distance coordinate of the focus of the parabola was 5.72 in the accuracy analysis and 6.6 in the reaction time analysis: both values are slightly above the midpoint of the Bubblesort range.

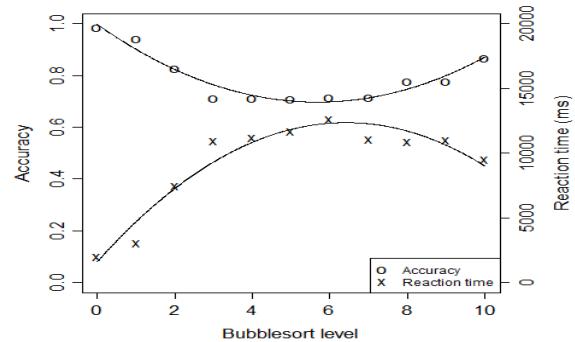


Figure 3: Experiment 2: Mean Accuracy and Mean Reaction Time vs. Bubblesort level (human subjects experiment).

General Discussion

We have presented evidence for an interactive activation model, NGRAMSWELL, of word reading with position-independent multi-scale feature detection. The model made qualitatively appropriate predictions about two new results in anagram solution: (1) manipulating the competition measure, JComp, which detects the intensity of jockeying among large-scale partial aggregations of letters, had a significant in-

fluence on accuracy and reaction times and (2) accuracy and reaction times vary quadratically with Bubblesort distance.

Result (1) extends prior work on anagram solution by identifying a new measure which may be an effective predictor of anagram solution times. NGRAMSWELL suggests that the predictive power of JCComp may stem from the self-organizing nature of the process of word detection. Self-organizing occurs when interactions among many autonomously acting but interacting elements give rise to structure at the scale of the group. One of the novel predictions of the self-organization approach to cognition is that structures at an intermediate scale between atomic micro-elements and fully well-formed objects play a significant role in making or breaking each perception episode (Tabor, Galantucci, & Richardson, 2004). The current results suggest that the n-grams with n just short of the length of the target are such intermediate structures in the anagram task.

Result (1) is also closely related to work on the effect of neighborhood size and type on word recognition under normal circumstances. Indeed, prior findings suggest that exchange of position of two adjacent letters produces less mental distortion than the replacement of one letter by a different one (e.g., Perea & Carreiras, 2006), even though the standard definition of lexical neighborhoods treats the exchange case as a more radical shift. The model proposed here suggests a way of organizing the encoding of mental structure that achieves an appropriate degree of positional fluidity. Although it is not a learning model, it may help guide the discovery of an appropriate learning model.

The Introduction suggested that, by examining how people solve anagrams, we might learn something about how they can sometimes perform well in a domain with which they've had little direct experience. Result (2) offers a specific insight about this issue: we can ask, why it is that people can read backwards fairly easily, compared to solving anagrams with Bubblesort levels in the mid-range (4-6). The answer seems to be that reversing letter order preserves the higher order relational structure of a sequence even though it destroys many local perceptual cues. It does this in virtue of a symmetry that is inherent in the nature of sequences in general. Perhaps what people are doing when they are successfully navigating untrodden territory is tuning into such universal regularities. This may be true even in the case of solving the hard anagrams (Bubblesort levels in the middle range), which people and NGRAMSWELL sometimes manage to solve by pop-out. We suggest exploring the deployment of symmetries in the parameter spaces of dynamical models to try to find out what conditions create the possibility of rich extrapolative generalization.

Acknowledgements

This research was supported in part by an NIH grant to Haskins Laboratories: HD40353.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al. (2007). The english lexicon project. *Behavior Research Methods*, 39, 445–459.

Dominowski, R. L. (1966). Anagram solving as a function of letter moves. *Journal of Verbal Learning and Verbal Behavior*, 5, 107–111.

Gilhooly, K. J., & Johnson, C. E. (1978). Effects of solution word attributes on anagram difficulty: a regression analysis. *Quarterly Journal of Experimental Psychology*, 30, 57–70.

Grimes, D., & Mozer, M. C. (2001). The interplay of symbolic and subsymbolic processes in anagram problem solving. In T. K. Leen, T. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems 13* (pp. 17–23). Cambridge, MA: MIT Press.

Knuth, D. E. (1973). The dangers of computer science theory. *Logic, Methodology and Philosophy of Science*.

Mayzner, M. S., & Tresselt, M. E. (1958). Anagram solution times: A function of letter and word frequency. *Journal of Experimental Psychology*, 56, 376–379.

Mayzner, M. S., & Tresselt, M. E. (1959). Anagram solution times: A function of transitional probabilities. *Journal of Experimental Psychology*, 63, 510–513.

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception, part 1: An account of basic findings. *Psychological Review*, 88(5), 375–402.

Mendelsohn, G. A. (1976). An hypothesis approach to the solution of anagrams. *Memory and Cognition*, 4(637–642).

Mendelsohn, G. A., & O'Brien, A. T. (1974). The solution of anagrams. a reexamination of the effects of transition letter probabilities, letter moves, and word frequency on anagram difficulty. *Memory and Cognition*, 2, 566-574.

Novick, L. R., & Cote, N. (1992). The nature of expertise in anagram solution. In *Proceedings from the fourteenth annual conference of the cognitive science society* (pp. 450–455). Hillsdale, NJ: Lawrence Erlbaum Associates.

Novick, L. R., & Sherman, S. J. (2004). Type-based bigram frequencies for five-letter words. *Behavior Research Methods, Instruments, and Computers*, 36, 397–401.

Novick, L. R., & Sherman, S. J. (2008). The effects of superficial and structural information on online problem solving for good versus poor anagram solvers. *The Quarterly Journal of Experimental Psychology*, 61(7), 1098–1120.

Perea, M., & Carreiras, M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychonomic Bulletin and Review*, 13, 418–422.

Tabor, W. (2003). Learning exponential state growth languages by hill climbing. *IEEE Transactions on Neural Networks*, 14(2), 444-446.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4), 355-370.