Balancing Long-Term Reinforcement and Short-Term Inhibition

Christian Lebiere (cl@cmu.edu)
Department of Psychology, Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15208 USA

Bradley J. Best (bjbest@adcogsys.com)
Adaptive Cognitive Systems
1709 Alpine Ave, Boulder, CO 80304 USA

Abstract

The ability to imperfectly but robustly enumerate a set of alternatives manifests itself in many human activities. However, many cognitive models have fundamental difficulties with this task, which often leads to degenerate behavior. The primary source of this problem is the conflict between mechanisms of long-term reinforcement and the need for short-term inhibition of recent items. Our analysis of a pair of pervasive domains of human activity finds that the long-term reinforcement process is balanced by a short-term inhibition. We have implemented this empirical finding in a variation of the knowledge reinforcement equation of the ACT-R architecture. This new mechanism not only prevents degenerate behavior in memory retrieval, but also emerges as a source of the power law distribution observed in the environment, supporting the proposition that the power law distribution arises from the interaction of the environment and cognition itself.

Keywords: Cognitive Architectures, Bayesian Learning, Human Memory, Power Law of Learning.

Introduction

A key aspect of many cognitive tasks, including planning and game playing, involves sequentially examining and evaluating a number of possible options. Many everyday tasks also have that characteristic, such as attempting to name all fifty states of the United States, or all members of one's research group.

Such enumeration is seldom exhaustive and unstructured. For instance, even (or especially) chess grandmasters seldom evaluate more than a few potential moves in any given position. Also, people's ability to enumerate large sets of objects also tends to be limited to a small number of items before repetition sets in, unless they resort to domain-specific strategies (e.g., relying on geographical or alphabetical order to name the 50 states).

However, as limited as this enumeration capacity might be, it also turns out to be highly pervasive and problematic for many current computational models of cognition. Many modern cognitive architectures and frameworks, unlike earlier models such as production systems (e.g., McDermott & Forgy, 1978), do not include explicit control mechanisms such as refraction, a mechanism that temporarily prevents the reapplication of a just-performed action, such as retrieving a specific piece of information. Even more problematic, some of those architectures, e.g., ACT-R (Anderson & Lebiere, 1998; Anderson et al., 2004), include mechanisms that reinforce a piece of information after it has been accessed, making it even more likely to be accessed in the future. Young (2003) has pointed out that the lack of refraction, together with those reinforcement mechanisms, can lead to pathological behaviors such as out-of-control looping where one piece of knowledge becomes so active that it is constantly retrieved at the expense of others.

This problem has also occurred in even more serious form in the development of higher-level languages that include implicit or explicit retrieval loops for checking logical conditions often involving universal quantifiers (Jones, Crossman, Lebiere, & Best, 2006; Jones, Lebiere, & Crossman, 2007). Nested loops are particularly difficult to deal with because they invalidate many of the potential attempts to deal with the problem. Of course, this kind of processing is cognitively very difficult, and people certainly don't perform it perfectly or completely, but they usually manage to accomplish the task to some extent while avoiding the pathological behavior often exhibited by the cognitive models. While modelers can often prevent such behavior by carefully crafting their models to imbed clever strategies reflecting their meta-knowledge of the task, such knowledge engineering methods lack cognitive plausibility. They also contribute to the brittleness and lack of portability of cognitive models when they are transferred to real world situations that lack the predictability, static nature, constrained scale, and limited complexity of laboratory experiments.

What is needed is an understanding of the human ability to enumerate a limited set of alternatives. Specifically, we need to analyze the environmental, cognitive and neural constraints that allow us to balance the long-term desire to reinforce more commonly accessed items to facilitate future retrievals together with the short-term need to avoid iterated retrievals that lead to looping behavior. These constraints, of course, must include those that come from observing human behavior. The constraints then need to be implemented in computational form that integrates with the other architectural mechanisms to provide the proper functional characteristics that allow the cognitive models to display the same robustness and adaptivity, as is commonly exhibited by human experimental participants, without recourse to constant engineering on the part of the modeler.

Environment Analysis

Following the rational analysis approach (Anderson, 1990), we view the environment as the primary shaper of cognition. Indeed, it was the analysis of Anderson & Schooler (1991) that led to the reinforcement mechanism in ACT-R that is largely responsible for the degenerate behavior described above. The primary insight of that analysis was that the odds of an item appearing in many common contexts such as email correspondence or newspaper headlines increased as a power law of frequency and decreased as a power law of recency. The latter, in particular, provides a boost in activation after an item is accessed, leading to a much higher probability of immediate subsequent retrieval unless that item is specifically excluded, such as through explicitly marking conditions on retrievals or tagging by a mechanism such as a declarative version of visual finsts (Pylyshyn, 1989). It is fair to ask what we might add to the prior analysis of Anderson and Schooler. Significantly, that analysis used domains that were not directly relevant to (or specifically excluded data related to) the short time scales of interest here. Therefore, we will follow the same rational analysis methodology in analyzing domains involving the consecutive access of large numbers of items, but rather than exclude items processed at short intervals, we will focus on them. We have selected two such domains that are pervasive in human activity and provide large amounts of regular data for such an analysis: language processing and arithmetic computations.

Language

To be able to draw broad and robust conclusions, we analyzed a broad array of text corpuses, from a scientific book chapter and proposal (about 30K words each) to classic books such as Dickens' "A Christmas Carol" (about 100K words) and Joyce's "Ulysses" (about 300K words) and a large section of a reference book, the Encyclopedia Britannica (about 300K words). Figure 1 plots the odds of a given word appearing a certain number of words after its previous appearance (a.k.a. lag), averaged over all words of each text. The plot is on a log-log scale in order to easily display the expected power-law relations.
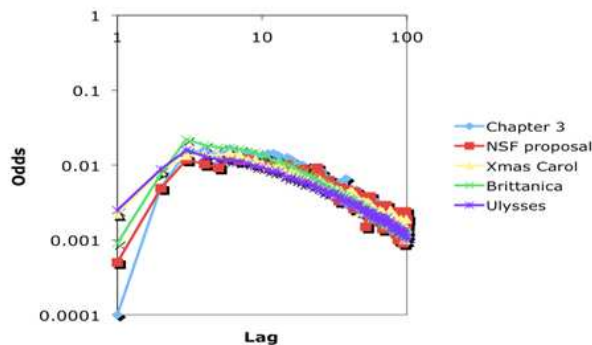
As observed, the relation is largely invariant of corpus, with statistically insignificant variations for very short lags (especially 1, which reflects infrequent word repetitions). For lags of 10 or more, the curve is roughly linear, i.e. the odds decrease as a power law of lag as in the original analysis. However, for short lags, the odds of appearance gradually deviate from the power-law, with a precipitous decline for immediate word repetitions (i.e. lag=1). This deviation can be interpreted as an inhibition of return in the environment that should be reflected in the appropriate cognitive mechanisms, inasmuch as it is the opposite dynamic to the current short-term reinforcement.
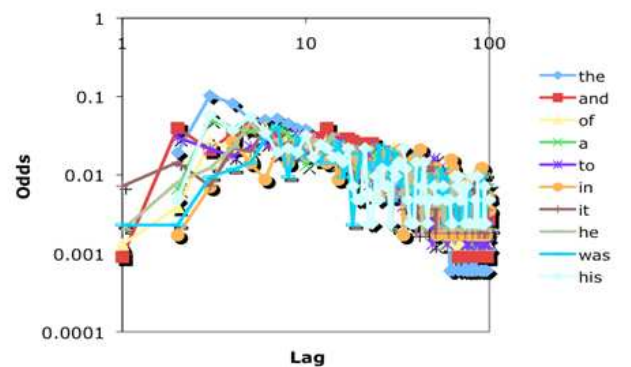

Figure 2: Odds of 10 most common words

We performed a more detailed analysis to rule out the possibility that this result was an artifact of grammatical rules or irregular words. Figure 2 establishes that the pattern holds for the ten most common words, albeit with additional noise resulting from the smaller sample size. Those words should probably be excluded because their syntactic roles are usually distinctive enough to result in procedural rather than declarative processing (if they are processed at all – articles are often not lexically processed). An important question is whether this pattern holds for less common words, and whether the period of inhibition is sensitive to word frequency. To answer that question, we categorized the words by frequency of appearance.

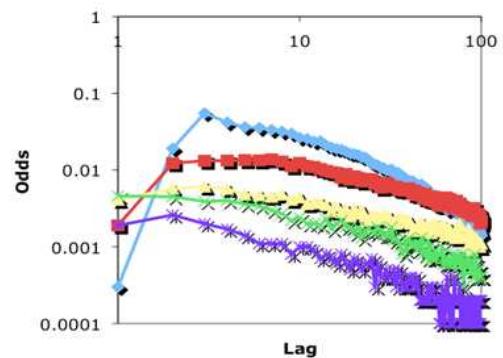Figure 1: Odds of appearance as a function of lag

Figure 3: Odds by word frequency in Ulysses

Figure 3 and 4 present a frequency analysis where words were grouped into five quintiles according to their frequency in the two largest texts. The curves for each quintile are parallel, capturing the frequency effects (other than for the first quintile of most common words at long lags)[1]. Both figures, but especially figure 4 (and one would expect Britannica to be more regular and less linguistically idiosyncratic than Ulysses), establish that the inhibition effect is not dependent on word frequency. Rather, the maximum odds of appearance occur around lag of 8-10 for all word frequencies. This is an essential piece of data in constraining the implementation of this environmental constraint in an architectural mechanism.
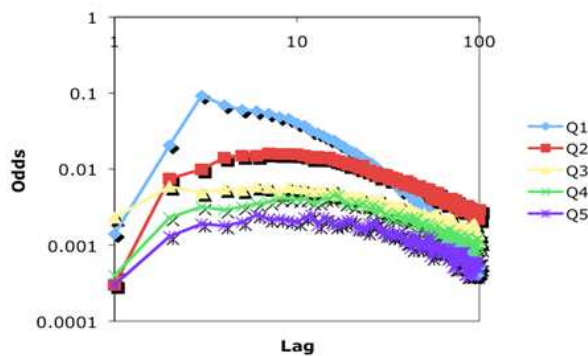

Figure 4: Odds by word frequency in Britannica

Arithmetic

Just as the original analysis by Anderson and Schooler had demonstrated the long-term statistical patterns over a range of domains, it is essential for us to show that the short-term inhibition trends found in the language corpuses are also present in other environments. As a second domain, we have chosen arithmetic facts for a number of reasons: it is a regular domain that like language is shaped by both the everyday world and formal schooling; it is a semi-formal domain that is relatively well-understood; and it is a domain that we have studied and modeled ourselves (Lebiere, 1998). Arithmetic knowledge, for example, is decomposed in a hierarchy of types, starting with counting facts and on to addition and multiplication. Each subsequent level of the hierarchy is taught sequentially (i.e., children first learn to count, then add, then multiply...) and depends upon the previous ones for reconstructing the current ones (i.e., addition by repeated counting, multiplication by repeated addition, etc). Unfortunately, knowledge of the actual statistical distributions of access to arithmetic knowledge is not directly available in the same form as the language corpuses. Textbooks provide one source of knowledge, but this source is relatively incomplete. Instead, we decided to

rely upon our validated cognitive model of the lifetime learning of arithmetic facts, and in particular its assumptions about the distribution and frequency of arithmetic facts. To avoid being overly dependent upon those assumptions, however, we studied in particular the statistical requests generated for one type of fact (counting) by another type (addition). This allowed us to reflect both the statistics (in terms of the distribution of addition problems) and the structure of the domain in terms of the requests for counting facts generated in solving addition problems by backup computation. Figure 5 displays the odds of a counting fact being accessed as a function of the lag since the previous access:
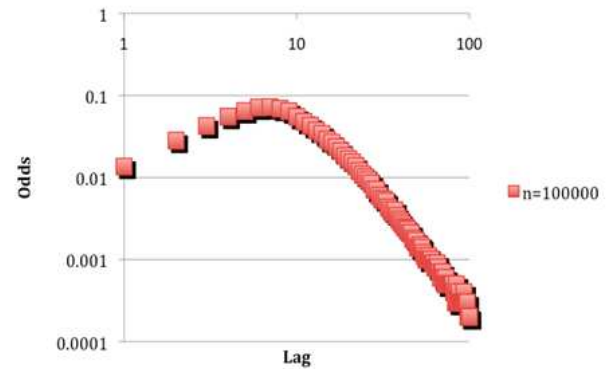

Figure 5: Odds of access to counting fact as a function of lag in addition by repeated counting problems

All the same patterns evident in Figure 1 are present: an initial inhibition for short lags, maximum odds slightly short of a lag of about 8-10, then the typical power law decay beyond that. As for the language corpuses, we decided to break down these aggregate odds into subcategories depending upon the frequencies of the facts accessed. Figure 6 presents the odds for five quintiles of decreasing frequencies.
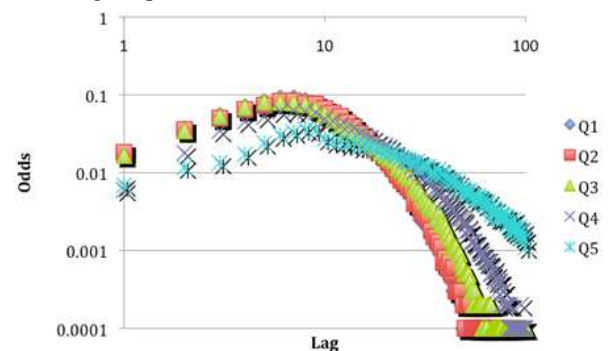

Figure 6: Odds of counting facts by quintiles

Again, the same patterns as in Figure 3 and 4 are present here as well: the curves for each quintile have the same characteristics as the aggregate curves, the maximum odds are reached at the same lags, and the curves are roughly parallel except for one. In the language corpuses, the most frequent quintile was the one exhibiting a different pattern

---

[1] This reflects, as discussed previously, the special syntactic role of the most common words that makes them highly likely to appear regularly and can be ignored for our purposes.

(a more steeply diving tail) because of the difficulty to produce text at long lags without mentioning very common words (e.g., articles). In the arithmetic corpus, the curve for the least frequent counting facts is the one exhibiting a different pattern (a flattening tail) because a few rare, large counting facts repeat at significantly longer lags since they only occur in a few, infrequent addition facts. For instance, in the 10x10 addition table, the counting fact stating that 18 follows 17 will only be needed for the most rare addition fact, 9+9, since the main statistical effect on fact distribution is that frequency decreases with size.

Computational Implementation

We now turn to implementing those findings in a cognitive architecture and combine the resulting mechanism with other, existing mechanisms to determine their interaction. We selected the ACT-R cognitive architecture for this implementation, although the resulting mechanism should be applicable to other architectures. As mentioned previously, ACT-R currently uses a number of techniques to prevent out-of-control retrieval loops resulting from short-term boosts in activation. A full discussion of those techniques is beyond the scope of this paper, but regardless it seemed that the solution to the problem should lay in the same location as its source, namely in the equation derived by Anderson & Schooler (1991) that describes the learning of the activation value of memory chunks as a function of history, specifically:

$$B_i = \log \sum_{j=1}^{n} t_j^{-d} \qquad \textbf{Base-Level Learning Equation}$$

The summation over all past n references to a chunk i, implement the power law of practice while the decay of each reference j as a function of time since reference t_j and decay rate d implements power law decay. However, as our analysis shows, those effects are only valid over longer time intervals, as exemplified by the linear trend in the log-scale graph of odds of appearance as a function of lag since the last appearance (e.g., see Figure 1 and 5). In the short run, within a lag of 5 to 10 appearances, there is a substantial decrease in odds of appearance that could be captured by a short-term inhibition mechanism. This effect is precisely what we are looking for to prevent out-of-control reinforcement where retrieval leads to a large short-term activation boost, which in turn leads to more retrieval, and so on until the system ends up completely deadlocked. A short-term inhibition component in the base-level activation learning equation would actually result in lower activation for the next retrieval(s), which would prevent the same chunk from being retrieved again and give other chunks a chance of being retrieved instead. This is exactly the functionality needed to implement many key patterns of behavior (e.g. Jones, Lebiere & Crossman, 2007) such as the ability of evaluating a set of alternatives.

We experimented with a number of different possible variations of the base-level learning equation, including power-law and exponential decay for the short-term inhibition component as well as different ways of combining the traditional reinforcement and the new inhibition factors. We have found the following form to be the most suitable in terms of both its functional properties and its ability to reproduce the data:

$$B_i = \log \sum_{j=1}^{n} t_j^{-d} - \log\left(1 + \left(\frac{t_n}{t_s}\right)^{-d_s}\right) \textbf{New BLL Equation}$$

The "1+" component of the new term is necessary to make sure that it has a strictly inhibitory effect, i.e. the sum is also greater than 1 meaning that the log is always positive. One can see that the new inhibitory term added is similar to the reinforcement term in being a power-law decaying term. However, this term only takes into account the most recent reference. Two new parameters have been added: a short-term decay rate d_s and a time scaling parameter t_s. Figure 7 displays the effect of this new term on the base-level activation and the effect of the new parameters. The first curve (BLL) displays the standard base-level curve, linear in log-log space. The second curve (PL(0.75;10)) displays the new base-level learning with a short-term decay rate of 0.75 (compared to the standard decay rate of 0.5) and a time-scaling parameter of 10. The proper effect seems to be generated but too weakly. The third curve (PL(1.0;10)) shows the effect of increased decay on the short-term inhibition, with a probability of retrieval at lag 1 roughly equal to that at lag 100, and a peak around lag 10, as generally observed in the data of Figure 1. The third curve (PL(1.0;5)) shows that decreasing the time scaling parameter from 10 to 5 effectively moves that peak from around lag 10 to around lag 5. It seems that the parameters of the second curve (short-term decay of 1.0 and time scaling constant of 10) are about right. The final two curves (PL(3;1.0;10) and PL(2;1.0;10)) show the effect of assuming multiple past references (3 and 2 times as many, respectively) as the original curve. One can see that the power law of practice is maintained at long lags (significantly above 10) while the difference is significantly reduced at short lags. The peak lag also seems to increase slightly with practice. All of these effects reproduce the data quite well (see previous figures).
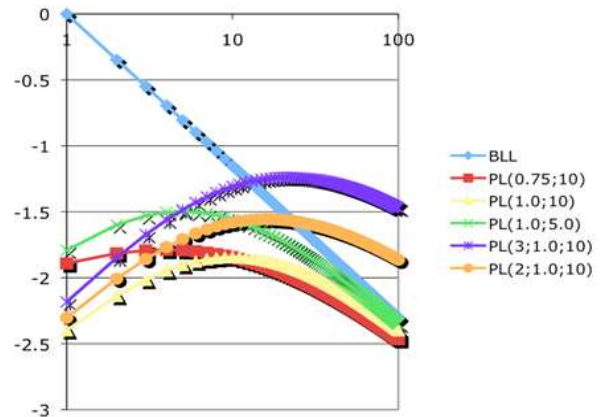

Figure 7: Inhibition effect of new base-level learning

This new equation allows for multiple interpretations. The main issue is identifying a neural construct this term can be associated with. The most direct interpretation would be to associate it with short-term depotentiation known to inhibit a neuron for a short period of time after its firing. Under that view, the short-term decay rate is simply another instance on a very small time scale of past proposals to introduce multiple decay rates in long-term memory corresponding to different time scales (Anderson, Fincham & Douglass, 1999), albeit here of an inhibitory nature. But another interpretation would be to associate this short-term inhibition of recently accessed information with its presence in short-term memory rather than its access in long-term memory. Under that interpretation, what is preventing the chunk from being retrieved from long-term memory is not that it couldn't be accessed there, but instead that it is still active in short-term memory, where access is much quicker and can therefore preempt the process of retrieval from long-term memory. The short-term decay rate is therefore the rate of decay in short-term memory, which is usually assumed to be much faster than that of long-term memory (e.g., Ridderinkhof et al., 2004).

Behavior

The next question is to determine the effectiveness of this mechanism in generating appropriate behavior as a part of the overall architecture. In a simple model of free recall where a number of chunks compete for retrieval without any constraint, this new equation has been effective at preventing out-of-control looping and implementing a functional, if probabilistic, version of round-robin access. This capability for declarative retrieval is similar to that exhibited by a model of procedural selection that learned production rule utilities to implement a flexible and probabilistic monitoring loop (Lebiere et al, 2008). Gradual, seamless transition from declarative to procedural control is a hallmark of human cognition. An important benefit of this new form of the base-level learning equation is to provide the activation calculus in declarative retrieval with the same sort of inhibition available to procedural selection through error-correction feedback in production rule utilities. However, it is relatively easy to show robustness (or at least non-pathological behavior) under careful control conditions as implemented by a combination of the task model and its environment. What we are looking for is greater robustness that prevents degenerate behavior even when very little or no control is exerted upon the model behavior by the control structure provided by the modeler or by the environment itself (including other agents). That is, we want an emergent robustness that does not originate from the model control structure authored by the modeler or that depends on triggers present in the environment (e.g. in experimental design) but instead that results directly from the cognitive architecture and its mechanisms. Having this kind of safeguard in the architecture would greatly relieve the burden of the cognitive modelers and allow the development of larger, smarter and more complex models capable of unanticipated (reasonable) behavior.

We therefore put the new base-level learning mechanism to a stringent test of robustness. Since the problem is out-of-control retrieval loops, and the traditional solution is a combination of chunk tagging (or finsts) with corresponding constraints on retrieval, we focused on a model of free recall (unconstrained by the environment). The model is given a fixed set of chunks and asked to retrieve any chunk without any constraints repeatedly (for a large number of times). Each chunk starts with the same initial history, and hence the same activation. The current base-level learning mechanism would immediately lead to a loop in which the same chunk would be retrieved over and over again while all the other ones would be forgotten. With our new base-level mechanism, however, we observed that a robust and flexible process of round-robin retrieval emerged. Figure 8 displays the frequencies of free recall of each chunk as a function of each chunk's rank order of retrieval frequency and the total number of recall iterations (varying from 100 to 1000 for 10 chunks).
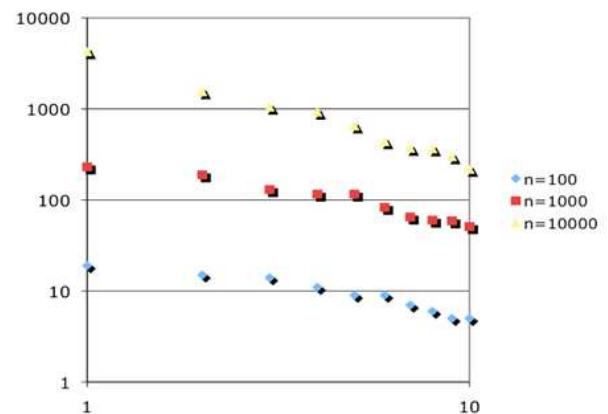

Figure 8: Frequencies of recall as function of item rank

The results are quite intriguing. The free recall behavior is neither the traditional degenerate winner-take-all nor its opposite, which would be an even distribution of recall frequencies over all chunks. This second extreme, a kind of perfect round-robin, would be cognitively implausible, as subjects in free recall experiments do forget some items and retrieve some multiple times. It might also be less than perfectly functional, since some items might deserve more frequent retrieval as a function of their characteristics such as urgency or rate of change (e.g. Lebiere et al, 2008). Instead, what emerges is a distribution of frequencies that slowly grows more uneven as the number of free recall iterations gets larger. Stronger items slowly emerge as some are gradually forgotten. What is more remarkable is that the frequency distribution seems to follow the power law observed in many natural environments, which is at the core of the base-level learning equation itself. While this may seem natural, it was not at all a given since, for example, the previous base-level learning equation that also reflected those distributions did not give rise to them

but instead led to degenerate, extreme distributions. Instead, finding one possible origin of this phenomenon in the cognitive system itself might provide one important piece of puzzle that has fascinated cognitive scientists for decades (e.g. Simon, 1955) and is increasingly viewed as resulting from an interaction of the environment and human cognition itself (e.g., Manin, 2008).

Conclusion

It is important to note the limitations of these results. First and foremost, full validation of this new mechanism requires showing that it can account directly for precise behavioral data, which we intend to do by taking existing models of memory phenomena and removing the safeguards handcrafted into them by modelers to prevent degenerate behaviors, and showing that previous results are preserved. Also, our rational analysis approach does not claim that this inhibition mechanism created these environmental effects, rather that our cognitive mechanisms have adapted to them. However, while other mechanisms (e.g. application of grammatical and arithmetic rules) might also be responsible for our overall cognitive performance in these tasks, the new inhibition mechanism definitely facilitates that performance. Finally, one can wonder whether those external task characteristics (e.g. grammatical rules against repetition, choice of base 10 arithmetic) might not have themselves evolved to reflect pre-existing cognitive characteristics. Such a determination would require modeling alternative task variants as well as performing the same analysis in more naturalistic domains.

We have presented data that demonstrate the reduced short-term likelihood of recent items appearing again in the environment and an extension to the ACT-R architecture that reflects this likelihood and produces more robust behavior than it did previously. We are currently exploring a number of additional lines of study. Empirically, we aim to replicate our findings in other human environments such as computer navigation, web environments and spatial navigation. Neurally, we aim to find correlates of inhibition in mechanisms such as short-term depotentiation. Behaviorally, we want to understand better how this new mechanism interacts with the rest of the architecture and which other emergent effects it might give rise to, such as the spacing effect. This approach underlies our belief that applying multiple constraints to architectural mechanisms is the best way to satisfy both scientific goals such understanding human behavior and practical goals such as developing more robust cognitive models.

References

Anderson, J. R. (1990). The Adaptive Character of Thought. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Fincham, J. M. & Douglass, S. (1999). Practice and retention: A unifying analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 1120-1136.

Anderson, J. R., & Lebiere, C. (1998). The Atomic Components of Thought. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological Science, 2, 396-408.

Anderson, J. R. & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving and F. I. M. Craik (Eds.) Handbook of memory, 557-570. New York: Oxford University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y . (2004). An integrated theory of the mind. Psychological Review 111, (4). 1036-1060.

Huberman, B. A., Pirolli, P., Pitkow, J. E. & Lukose, R. M. (1998). Strong Regularities in World Wide Web Surfing. Science 280 , 5360, 95-97.

Jones, R.M., Crossman, J. A., Lebiere, C., & Best, B. J. (2006). An abstract language for cognitive modeling. In Proceedings of the 7th International Conference on Cognitive Modeling. Trieste, Italy

Jones, R., Lebiere, C. & Crossman, J. A. (2007). Comparing Modeling Idioms in ACT-R and Soar. In Proceedings of the 8th International Conference on Cognitive Modeling. Ann Arbor, MI.

Lebiere, C. (1998). The dynamics of cognition: An ACT-R model of cognitive arithmetic. Ph.D. Dissertation. CMU Computer Science Dept Technical Report CMU-CS-98-186. Pittsburgh, PA. Available at http://reports-archive.adm.cs.cmu.edu/

Lebiere, C., Archer, R., Warwick, W., & Schunk, D. (2005) Integrating modeling and simulation into a general-purpose tool. In Proceedings of the 11th International Conference on Human-Computer Interaction. July 22-27, 2005. Las Vegas, NV.

Lebiere, C., Archer, R., Best, B., & Schunk, D. (2007). Modeling pilot performance with an integrated task network and cognitive architecture approach. In Foyle, D. & Hooey, B. (Eds.) Human Performance Modeling in Aviation. Mahwah, NJ: Erlbaum.

McDermott, J. & Forgy, C. (1978). Production system conflict resolution strategies. In D. A. Waterman & F. Hayes-Roth (Eds.), *Pattern-Directed Inference Systems*, 177-199. Academic Press.

Manin, D.Y. (2008). Zipf's law and avoidance of excessive synonymy. *Cognitive Science, 32(7)*, 1075-1098.

Pylyshyn, Z.W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition, 32, 65-97.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The Role of the Medial Frontal Cortex in Cognitive Control. Science, Vol. 306. no. 5695, pp. 443 – 447.

Simon, H. A. (1955). On a Class of Skew Distribution Functions. Biometrika 42 (3-4): 425–440.

Young, R. M. (2003). Should ACT-R include rule refraction? In Proceedings of the Tenth Annual ACT-R Workshop. Pittsburgh, PA.