

Connections between ACT-R's declarative memory system and Minerva2

Cvetomir M. Dimov (cvetomir.dimov@unil.ch)

Department of Information Systems, University of Lausanne
CH-1015, Switzerland

Abstract

As a first step towards applying ACT-R to problems of likelihood judgment, we draw parallels between ACT-R and HyGene. More specifically, in the spirit of theory integration, we demonstrate the relation between ACT-R's declarative memory system and the core of HyGene: Minerva2. We first start by transforming ACT-R's activation equations into what is in our view a more intuitive form. This form then allows us to more transparently see the correspondence of the effect of prior history between the two theories and of the current context between them. The results provide insights into the workings of the two theories and open an avenue for future attempts of theory integration, not only between the two theories, but also to related theories of memory. Moreover, we hope these results will be important steps toward testing ACT-R's capabilities of accounting for judgment phenomena.

Keywords: ACT-R, HyGene, Minerva2, declarative memory

Introduction

The cognitive architecture ACT-R (Anderson, 2007) is arguably the most advanced integrated theory of cognition at the moment. A testament to its generality is the broad spectrum of phenomena to which it has been applied. These include phenomena as diverse as analogy making (Salvucci & Anderson, 2001), past tense learning (Taatgen & Anderson, 2002) and solving the Tower of Hanoi puzzle (Anderson & Douglass, 2001).

One set of phenomena to which ACT-R has been applied are those in the field of judgment and decision making. Specifically, the memory system of this cognitive architecture has been used to explore the applicability of various decision strategies (Marewski & Schooler, 2011) and to investigate the properties of the fluency heuristic (Schooler & Hertwig, 2005). In addition, the full power of the architecture was used to construct 39 implementations of the recognition heuristic (Marewski & Mehlhorn, 2011) and identify those that best describe behavioral data.

A problem in the domain of judgment and decision making to which ACT-R was never, to our knowledge, applied is how people make probability judgments. Probably the most complete psychological theory of probability judgment is HyGene (Thomas, Dougherty, Sprenger, & Harbison, 2008). Thus, a first step in applying ACT-R to problems of probability judgment would be to work towards the integration of ACT-R and HyGene by exploring the compatibility of the two theories.

HyGene extends MINERVA-DM (Dougherty, Gettys, & Ogden, 1999) by adding semantic memory and working memory storages. MINERVA-DM was first introduced to account for various likelihood judgment phenomena, such as frequency judgments and some judgment biases, and is itself based on the theory of memory Minerva2 (Hintzman, 1984).

Thus, many of HyGene's properties stem from the very properties of Minerva2.

In the spirit of theory integration (Mischel, 2009, criticized the lack of theory integration efforts in psychology), we will demonstrate the close correspondence between ACT-R's declarative memory and Minerva2. We will first briefly introduce ACT-R's declarative memory equations and Minerva2. We will then present what is in our mind a more intuitive functional form of activation - the exponent of activation. We will then derive the optimized learning equation in the exponentiated form and demonstrate the conditions under which it provides a good approximation to base-level activation. This will allow us to relate influence of prior exposure according to the two aforementioned theories of memory. We will then compare the effect of context on activation according to the two theories and again demonstrate the close correspondence between the two.

ACT-R's declarative memory system

ACT-R's declarative memory system integrates retrieval probability and retrieval time of a memory chunk through a single quantity: the memory activation of that chunk. Memory activation A_i of a chunk is modeled as a function of prior exposure to the object represented by the chunk (base-level activation), context relevance (spreading activation of chunk) and noise:

$$A_i = B_i + SA_i + \varepsilon = B_i + \sum_j W_j S_{ji} + \varepsilon, \quad (1)$$

In this equation, B_i is the base-level activation, which reflects the recency and frequency of practice of the chunk i . ε is the noise value, which consists of permanent noise and instantaneous noise computed at the time of a retrieval request. We will only consider instantaneous noise below. W_j is the amount of activation from source j , while S_{ji} is the strength of association from source j to chunk i .

The equation describing learning through base-level activation for chunk i is:

$$B_i = \ln \sum_{j=1}^n t_j^{-d} \quad (2)$$

where n is the number of presentations for a chunk i , t_j is the time since the j^{th} presentation, and d is the decay parameter. According to the base-level equation, each time an item is presented, its activation increases. Then, with time, its activation decays, described with a power function.

The strength of association S_{ji} between two chunks j and i is 0 if chunk j is not the value of a slot of chunk i (also,

j and i should not be the same chunk). If the two chunks have common slots, the association strength is set using the following equation:

$$S_{ji} = S - \ln(\text{fan}_j) \quad (3)$$

In this equation, S is the maximum associative strength and fan_j is the number of chunks in declarative memory in which j is the value of a slot. In addition, 1 is added for each chunk (in this case, chunk j) being associated with itself. Intuitively, the amount of spreading activation from a slot of the chunk currently in the focus of attention is distributed equally among all chunks with that slot in declarative memory.

A brief introduction to Minerva2

Minerva2 represents stimuli as feature-lists, which get stored as episodic traces in memory upon encoding. Each repetition produces a new trace of that item. Upon cued retrieval, all episodic traces in memory are contacted and compared to the cue, and produce a cumulative output in working memory called an echo. In the echo, each trace is weighted by its activation.

Activation $A(i)$ of an episodic trace in Minerva2 is a cubic function of its similarity to the cue. It is expected that the activation of a trace is close to 0 for a randomly chosen trace and it increases superlinearly to 1 when the cue and the trace are the same.

Exponentiating memory activation

We consider that for the current purposes the exponent of memory activation is a more intuitive quantity to work with than memory activation on its own:

$$A_i^e = e^{A_i} = e^{B_i + SA_i} = e^{B_i} e^{SA_i} = B_i^e SA_i^e \quad (4)$$

Base-level activation

In the following we will use the non-log-transformed version of base-level activation B_i^e (Anderson & Schooler, 1991; Anderson, Fincham, & Douglass, 1999), because it demonstrates the effect of prior exposure on activation more transparently:

$$B_i^e = \sum_{j=1}^n t_j^{-d} \quad (5)$$

Derivation of the optimized learning equation To derive the optimized learning equation, we suppose that we know the time since the chunk was created, called its lifetime L . Lets also suppose that it has been periodically strengthened n times since its creation with a period of $T = \frac{L}{n}$. Given these assumptions, we can derive a new form of the base-level activation:

$$B_i^e = T^{-d} \sum_{j=1}^n j^{-d} \quad (6)$$

We can approximate the sum above with an integral (Anderson et al., 1999), and arrive at the simplified equation:

$$B_i^e = T^{-d} \frac{n^{1-d}}{1-d} \quad (7)$$

At first sight, it seems that the base-level activation is a power function of n (Anderson et al., 1999). Note however that T is also a function of n . Taking this into account, we arrive at the final expression for B_i^e :

$$B_i^e = \frac{L^{-d}}{1-d} n \quad (8)$$

This is (the exponent of) the optimized learning equation, available, for example, in ACT-R's documentation. Note that it is a *linear function* of n , discounted only by the chunk's lifetime L .

Generality of the optimized learning equation Some might argue that equation 8 rests on the rather strict assumption of periodicity. To test how much of a restriction that is, lets start with the following "noisy" base-level activation equation:

$$B_i^n = \sum_{j=1}^n (jT + j^{1+d}\epsilon_j)^{-d} = \sum_{j=1}^n (jT)^{-d} \left(1 + \frac{j^d\epsilon_j}{T}\right)^{-d}, \quad (9)$$

where the noise j periods ago is $j^{1+d}\epsilon_j$.

Using the approximation $(1+a)^n \approx 1+na$, we can separate the noise term from the rest:

$$B_i^n = \sum_{j=1}^n (jT)^{-d} \left(1 - \frac{j^d d \epsilon_j}{T}\right) = B_i - dT^{-(1+d)} \sum_{j=1}^n \epsilon_j \quad (10)$$

If ϵ has an expectation of 0, then our noisy base-level activation equation will be unbiased. Notice that the noise that we added grows as a *superlinear* function of the number of periods j .

Spreading activation

According to ACT-R the slots of the chunk present in the imaginal buffer spread activation to chunks in declarative memory. The exponentiated spreading activation is equal to:

$$SA_i^e = e^{SA_i} = e^{\sum_j \frac{w_{im}}{n_{im}} S_{ji}} = \prod_j e^{\frac{w_{im}}{n_{im}} S_{ji}}, \quad (11)$$

where w_{im} is the total amount of spreading activation from the imaginal buffer and n_{im} is the number of slots in the imaginal buffer's chunk. Intuitively, the total amount of spreading activation that the imaginal buffer can provide is *equally* distributed among its slots.

What remains to unpack is S_{ji} , which is the difference between the maximum possible amount of spreading activation S and the number of chunks in declarative memory with that slot value fan_j . For simplicity we set $w_{im} = n_{im}$ so that these two cancel each other out. The contribution to spreading activation from a slot is:

$$e^{S_{ji}} = \frac{e^S}{\text{fan}_j}, \quad (12)$$

Note that the total spreading activation from a chunk slot in the imaginal buffer is *equally* distributed among all chunks in declarative memory which contain that slot (Anderson & Reder, 1999, discuss the general equation for spreading activation).

The effect of prior history on activation according to ACT-R and Minerva2

In this section we will use equation 8 to draw parallels between memory activation according to Minerva2 and ACT-R. We will consider only the effect of prior history. For Minerva2 this means a constant similarity S . For ACT-R will assume that there is constant spreading activation.

Minerva2 encodes each observation as a separate trace. When memory is probed, each trace in memory contributes with a certain activation to the echo intensity and content. Given a similarity, the activation of a trace is $A = S^3$. For an occurrence frequency of an event n , we will have a total activation of the n traces related to that event of $A^{tot} = nA = nS^3$. That is, the relative total activation of all traces related to two events is simply a function of their relative occurrence frequencies:

$$\frac{A_1^{tot}}{A_2^{tot}} = \frac{n_1 S^3}{n_2 S^3} = \frac{n_1}{n_2} \quad (13)$$

According to ACT-R, using equation 8, the relative activation of two chunks is:

$$\frac{B_1^e}{B_2^e} = \frac{L^{-d}(1-d)n_1}{L^{-d}(1-d)n_2} = \frac{n_1}{n_2} \quad (14)$$

It is important to emphasize that this result holds only for chunks with the same lifetime L . However, this is not necessarily a strong limitation to our result: In learning experiments often all pieces of information have the same lifetime. That is, learning usually starts at the same point - in the beginning of the experiment. And even if this is not the case, as demonstrated, there is some variability in the starting point allowed. In addition, it is often true for real-life memories such as general knowledge facts that they have similar lifetimes.

When there is considerable departure from this assumption, one can plug in the actual lifetime of memories to estimate their relative activation. For example, let us assume that one chunk of memory is 4 times older than another. That is, we assume that $L_1 = 4L_2$. Let us also take the standard value of the decay parameter $d = 0.5$. For these values the relative activation of the two chunks will be:

$$\frac{B_1^e}{B_2^e} = \frac{L_1^{-d}(1-d)n_1}{L_2^{-d}(1-d)n_2} = \frac{n_1}{2n_2} \quad (15)$$

This means that the relative activation of the younger chunk will be boosted two times compared to that of the older chunk. Generally, with the default decay parameter, if a chunk is r^2 times younger, its relative activation will be r times higher. This is a testable difference between ACT-R and Minerva2 and could be another source of bias, which Minerva2 does not account for.

The effect of context on activation according to ACT-R and Minerva2

For simplicity, let's assume that there are a total of m slots in the imaginal buffer's chunk and that $w_k = m$. Also, we

will assume that all slots have the same fan. If for chunk i a total of k slots spread activation, the total amount of spreading activation is:

$$SA_i^e = \prod_j e^{S_{ji}} = \left(\frac{e^S}{fan} \right)^k \quad (16)$$

Thus, spreading activation is power function of the number of matching slots. According to Minerva2, activation changes as the cube of the number of matching values:

$$A = S^3 = \left(\frac{k}{m} \right)^3 \quad (17)$$

Qualitatively, both expressions are superlinear functions of k , albeit with different functional forms. However, note that there is no strong justification for the latter expression. Its purpose is to allow "those items in memory that are most similar to the test probe to dominate the overall echo from secondary memory, while preserving the sign" (Dougherty et al., 1999, p. 183). ACT-R's spreading activation also affords this and is in fact consistent with other exemplar models such as Context Theory (Medin & Schaffer, 1978), where similarity of two stimuli follows a multiplicative rule. Such parallels can also be drawn between ACT-R and the more recent model of recognition memory REM (Shiffrin & Steyvers, 1997). In REM, the likelihood ratio λ_j plays the role of activation. Under the assumption that the values of a feature in a memory image V_{kj} are all equal, the expression for λ_j reduces to:

$$\lambda_j = (1-c)^{n_{jq}} \prod_{k \in M} \frac{c + (1-c)g(1-g)^{V_{jk}}}{g(1-g)^{V_{jk}}} \quad (18)$$

$$\lambda_j = (1-c)^{n_{jq}} \left(\frac{c + (1-c)g(1-g)^V}{g(1-g)^V} \right)^m, \quad (19)$$

where n_{jq} is the number of nonzero features, M are the indices of non-zero features that match, and m is the number of features that match. Generally, exemplar and memory models seem to converge to the notion that the relevance of a item of memory increases in a multiplicative fashion with the number of matching features.

Discussion and Conclusion

At first sight, ACT-R's declarative memory theory and Minerva2 seem quite different: ACT-R uses the notion of a memory chunk with slots, while Minerva2 uses vectors with binary values. In addition, ACT-R explicitly models memory decay, while according to Minerva2 the only way for information to be not perfectly recorded in memory is through imperfect encoding. Finally, ACT-R envisions a single internal representation of an external object, whose memory strength increases upon each encounter of that object or recall of the internal representation. Minerva2 on the other side saves separate traces upon each encounter.

Given these differences, it is not obvious at which points these two theories will be tangential to each other. Our results show that both as a function of prior history and current context the theories can make similar predictions. More

specifically, according to both theories, in many experimental conditions we expect relative activation to be a function only of relative presentation frequency. As for context, it rapidly increases the activation of memory traces similar to it. The correspondence between the functional form of this increase according to ACT-R and some exemplar models suggests an exploration of that very functional form as a viable alternative to Minerva2's current cubic function.

More importantly, the results presented above also indicate the plausibility of accounting for judgment phenomena with ACT-R. Future efforts should aim at fully implementing HyGene into ACT-R and repeating the very simulations that MINERVA-DM and HyGene have already presented. Of especial interest would be an exploration of those conditions, in which ACT-R's predictions differ from those of Minerva2, such as when chunks have different lifetimes or different fans.

We conclude with a wish for more frequent theory integration efforts in psychology, as we hope that the aforementioned results demonstrated.

Acknowledgments

We thank the Swiss National Science Foundation for generously funding Cvetomir Dimov through Project 146702.

References

- Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.
- Anderson, J. R., & Douglass, S. (2001). Tower of hanoi: evidence for the cost of goal retrieval. *Journal of experimental psychology: learning, memory, and cognition*, 27(6), 1331.
- Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: a unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1120.
- Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, 128(2), 186.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological science*, 2(6), 396–408.
- Dougherty, M. R., Gettys, C. F., & Ogden, E. E. (1999). Minerva-dm: A memory processes model for judgments of likelihood. *Psychological Review*, 106(1), 180.
- Hintzman, D. L. (1984). Minerva 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2), 96–101.
- Marewski, J. N., & Mehlhorn, K. (2011). Using the act-r architecture to specify 39 quantitative process models of decision making. *Judgment and Decision Making*, 6(6), 439–519.
- Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: an ecological model of strategy selection. *Psychological Review*, 118(3), 393.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, 85(3), 207.
- Mischel, W. (2009). The toothbrush problem. *Association for Psychological Science Observer*, 21(11), 1.
- Salvucci, D. D., & Anderson, J. R. (2001). Integrating analogical mapping and general problem solving: The path-mapping theory. *Cognitive Science*, 25(1), 67–110.
- Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological review*, 112(3), 610.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: Remretrieving effectively from memory. *Psychonomic Bulletin & Review*, 4(2), 145–166.
- Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say broke? a model of learning the past tense without feedback. *Cognition*, 86(2), 123–155.
- Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. (2008). Diagnostic hypothesis generation and human judgment. *Psychological review*, 115(1), 155.