



Institute of
Cognitive Science
CARLETON UNIVERSITY
Matthew A. Kelly
Kam Kwok
Robert L. West

Holographic Declarative Memory:

A Scalable Memory Module for ACT-R

- Holographic Declarative Memory (**HDM**) is a new module for ACT-R, implemented for Python ACT-R (Stewart & West, 2006).
- HDM is an alternative to ACT-R's Declarative Memory (**DM**).
- HDM replaces DM's symbols with *holographic vectors* (Plate, 1995) and implements a holographic theory of memory based on **DSHM** (Rutledge-Taylor, Kelly, West, & Pyke, 2014) and **BEAGLE** (Jones & Mewhort, 2007).

Holographic Declarative Memory

- Symbolic Architectures (e.g., ACT-R DM)
 - concepts and relations between concepts represented as text
 - lists of slot-value pairs
 - **“name:cat type:animal look:furry”**
- Vector Symbolic Architectures (e.g. HDM)
 - concepts and relations between concepts represented as vectors
 - **“name:cat type:animal look:furry”**
 - = [0.0916, 0.5175,-0.8271, 0.1983, 0.0197 ...]
- Connectionist / Neural Network Architectures
 - neurons and neural connectivity represented by vectors of patterns of activation and matrices of connection weights

Architectures

- Vectors can be understood as describing coordinates in a high-dimensional space
- Points close in space have similar meaning
- Allows for shades of meaning and partial matching

astronomy
physics
chemistry
psychology
biology
scientific
mathematics
technology
science
scientists
research

Vector space

- Holographic vectors retain the expressive power of symbols
 - Compactly store complicated, recursive relations between ideas
- Holographic vectors have a similarity metric, allowing for...
 - Shades of meaning and fuzzy / partial matching
 - Lossy compression for modeling forgetting
 - Fault tolerance
- HDM can enhance ACT-R's ability to:
 - Scale to big data / model long-term learning
 - Learn association strengths from experience
 - Provide a bridge to neural realization

Why use holographic vectors?

- Holographic models of memory in the literature
- Case Study: The Fan Effect
 - What is the fan effect?
 - The ACT-R DM model of the fan effect
 - The HDM model of the fan effect
- Results: How does DM and HDM compare?
- Analysis: Why does HDM work?
- Conclusions / future work

In what follows ...

- Explain and predict a variety of human **memory** phenomena
 - **Fan** effect (Rutledge-Taylor et al., 2014)
 - **Serial recall** and **free recall** of lists (Franklin & Mewhort, 2015)
 - **Implicit** learning (Jamieson & Mewhort, 2011)
- **Analogical reasoning**
 - (Plate, 2000; Eliasmith & Thagard, 2001)
- Simple **problem-solving** tasks
 - (Eliasmith, 2013; Rutledge-Taylor et al., 2014)
- **SPAUN**, the world's largest functional **brain model**
 - (Eliasmith, 2013)
- **BEAGLE**, a model of learning **word meaning** from a corpus
 - (Jones & Mewhort, 2007)

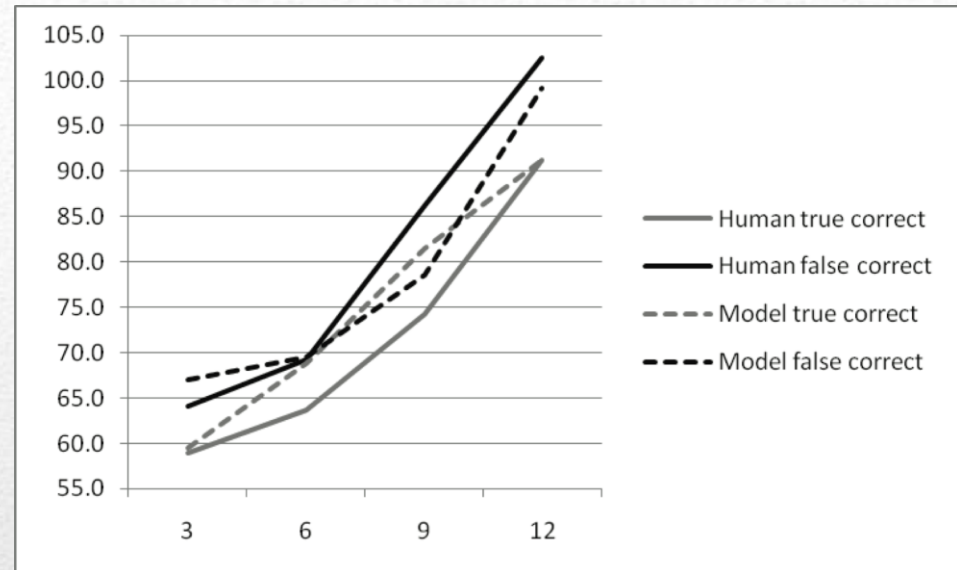
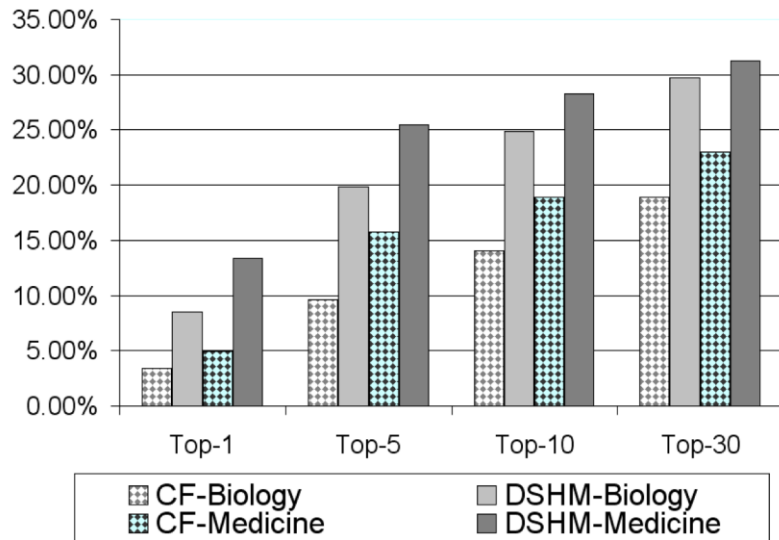
Holographic memory in the literature



- Takes a corpus as input
- Produces a set of vectors representing word meaning
- Similarities between vectors produce clusters of topic and part of speech
- Vector similarities predict semantic priming data

BEAGLE

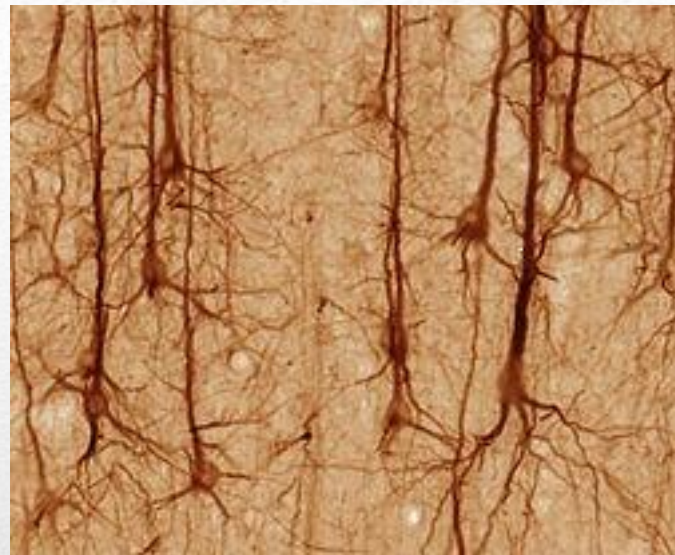
(Jones & Mewhort, 2007)



- Applies BEAGLE to non-linguistic stimuli
- Models two-term and three-term fan effect
- Models rock-paper-scissors play
- Effective recommender system for movies or research papers

DSHM (Rutledge-Taylor et al., 2014)

- **ACT-R** describes mental processes and **brain areas** associated with them, but does not address the question of how those mental processes are carried out at the neural level.
- **Holographic vectors** can be implemented in realistic neural models.
- **HDM** can be straightforwardly implemented in the **Neural Engineering Framework** (Eliasmith, 2013), a theory of neuro-computation.



Neural realization (Eliasmith, 2013)

- “the **hippy** is in the **park**”

- “the **hippy** is in the **bank**”

fan(**hippy**) = 3

- “the **hippy** is in the **store**”

fan(**store**) = 2

- “the **officer** is in the **store**”

fan(**officer**) = 1

Fan Effect (Anderson, 1974)

- participants are **slower** to recognize or reject sentences that contain concepts that have a **higher fan**.

- **availability** of information in memory with respect to a cue is related to the **probability** of that piece of information conditional on the cue.

fan(**hippy**) = 3

fan(**store**) = 2

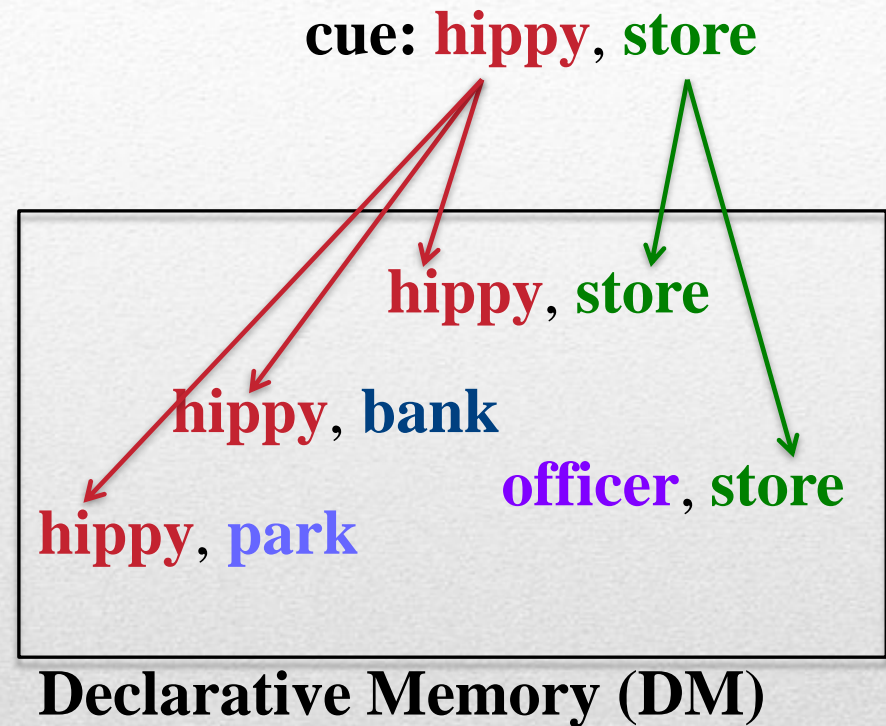
fan(**officer**) = 1

Fan Effect (Anderson, 1974)

- sentences are represented as *person, location* **chunks** in **DM**

- when the model is cued, **activation** spreads to chunks that share concepts with the **cue**

- DM retrieves most active chunk



ACT-R DM model

- Reaction time T is calculated as:

$$T = I + Fe^{-A_i}$$

- Activation A_i of chunk i is calculated as:

$$A_i = B_i + \sum_{j=1}^n W_j S_{ji}$$

- Association strength S_{ji} with concept j is:

$$S_{ji} = S + \ln(P(i/j))$$

- Where $P(i/j) = 1 / \text{fan of } j$

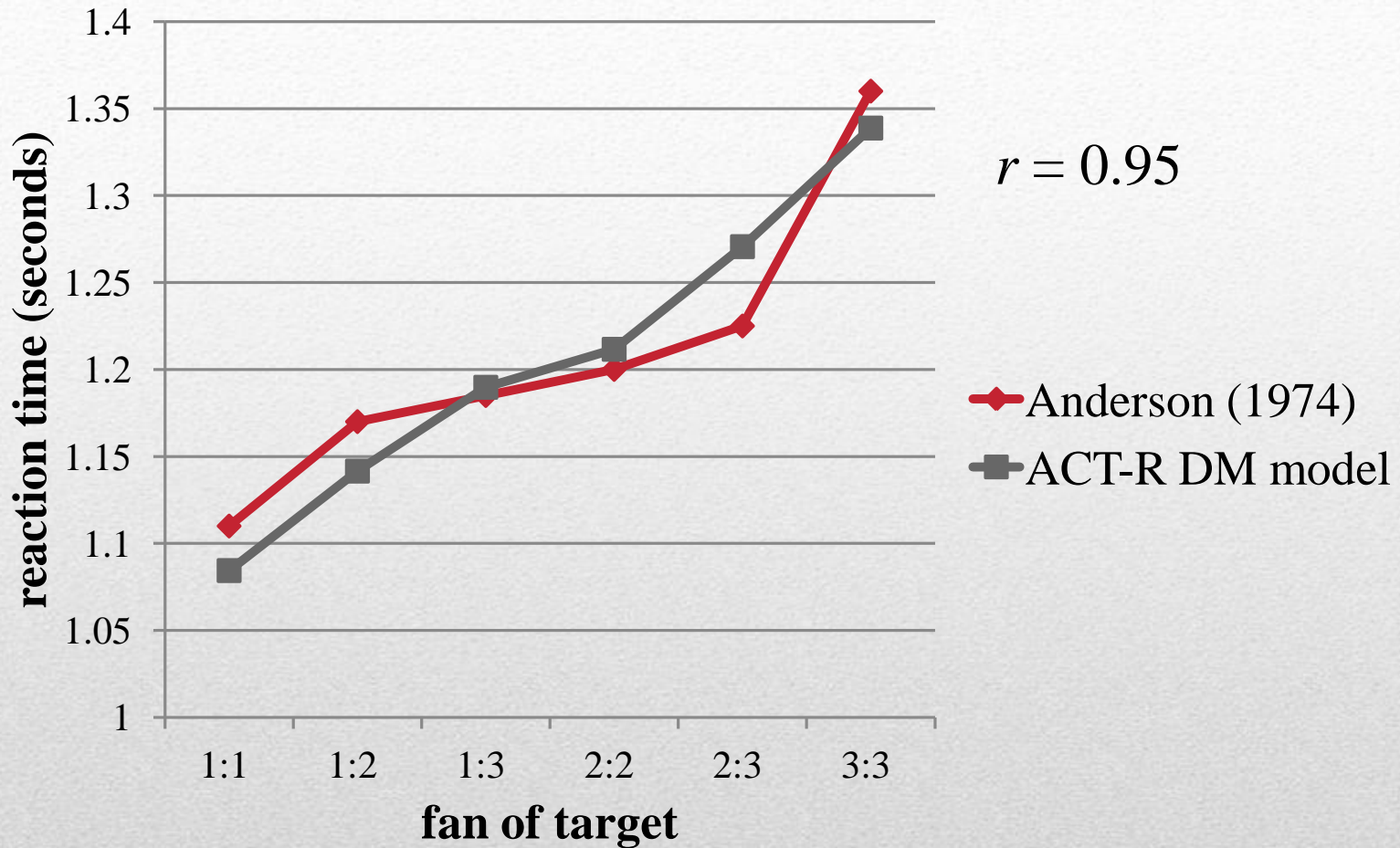
ACT-R DM model

- Anderson and Reder's (1999) model is, in milliseconds:

$$T = 233(f_{\text{person}} f_{\text{place}})^{1/3} + 845$$

- where f_{person} is the person's fan and f_{place} is the place's fan

ACT-R DM model
(Anderson & Reder, 1999)



ACT-R DM vs. Human

For each task symbol (or concept) there are two vectors:

environmental vector

- a random vector that stands for what the symbol looks like

memory vector

- a continuously updated vector of the symbol's associations

Additionally, there is one special vector used in all associations:

placeholder vector Φ

- can be read as ?, i.e., the value that we want to retrieve.
- acts as a stand-in for the purposes of storage and retrieval.

HDM model

* circular convolution is used to create associations between symbols in a sequence.

+ addition is used to add new associations to a memory vector

\mathbf{P}_{before} is a permutation indicating that the permuted vector comes earlier in a sequence.

\mathbf{P}_{slot} is a permutation indicating that the permuted vector is the value associated with the slot *slot*.

Holographic vectors

“the **hippy** is in the **park**”

$\mathbf{m}_{\text{hippy}}(\text{updated}) = \mathbf{m}_{\text{hippy}} + (\mathbf{P}_{\text{before}} \Phi) * \mathbf{e}_{\text{park}}$ “what came before **park**?”

$\mathbf{m}_{\text{park}}(\text{updated}) = \mathbf{m}_{\text{park}} + (\mathbf{P}_{\text{before}} \mathbf{e}_{\text{hippy}}) * \Phi$ “what came after **hippy**?”

To **encode** an association in HDM, memory vectors are updated with all **questions** to which the memory vector’s **concept** is an appropriate answer given HDM’s experiences.

Encoding associations

To test if “the **hippy** is in the **park**” is in memory, two cues are constructed:

$$\mathbf{q}_{\text{hippy?}} = (\mathbf{P}_{\text{before}} \mathbf{e}_{\text{hippy}}) * \Phi \quad \text{“what came after hippy?”}$$

$$\mathbf{q}_{\text{?park}} = (\mathbf{P}_{\text{before}} \Phi) * \mathbf{e}_{\text{park}} \quad \text{“what came before park?”}$$

The memory vectors most **similar** to these cues are retrieved.

Similarity is measured as the **cosine** of the angle between vectors.

Retrieving associations

Activation is calculated as the **mean** of the **cosines** between each cue and the memory vector substituted out to create the cue:

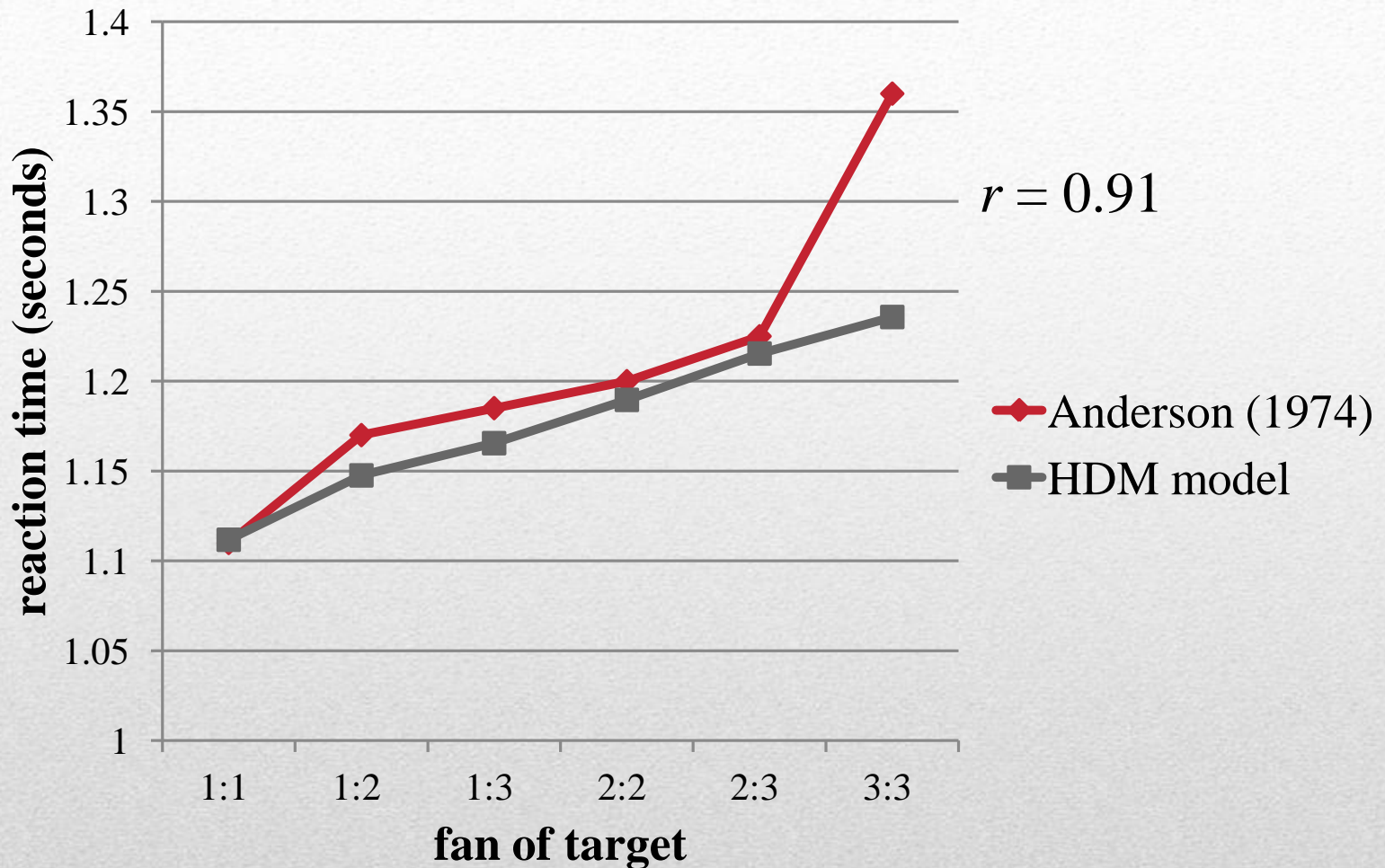
$$A = 0.5 \text{ cosine}(\mathbf{q}_{\text{hippy?}}, \mathbf{m}_{\text{park}}) + 0.5 \text{ cosine}(\mathbf{q}_{\text{?park}}, \mathbf{m}_{\text{hippy}})$$

Reaction **time** is computed using **DM**'s reaction time equation:

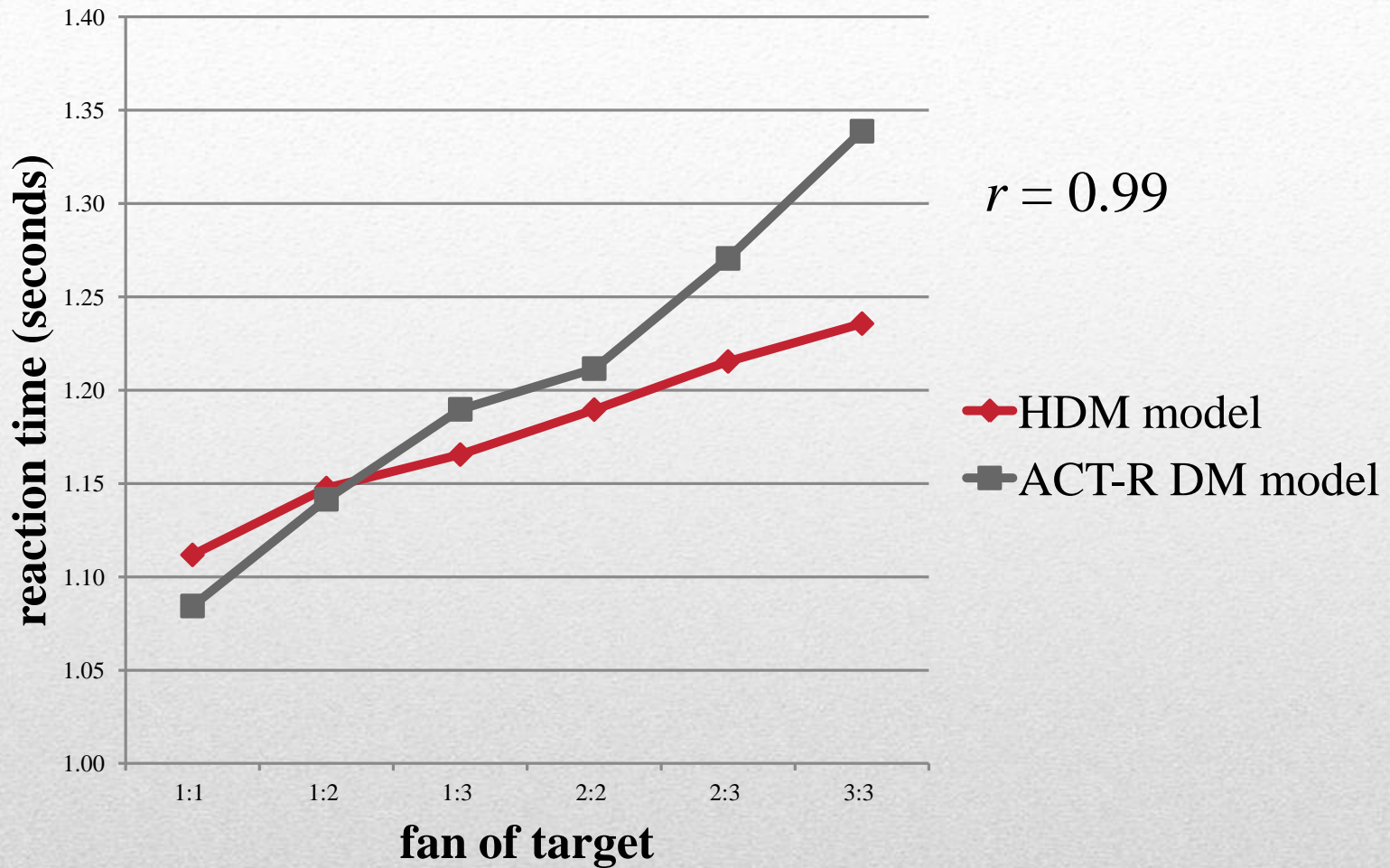
$$T = I + Fe^{-Ai}$$

Parameters I and F set to the values used by Anderson and Reder's (1999) ACT-R DM model.

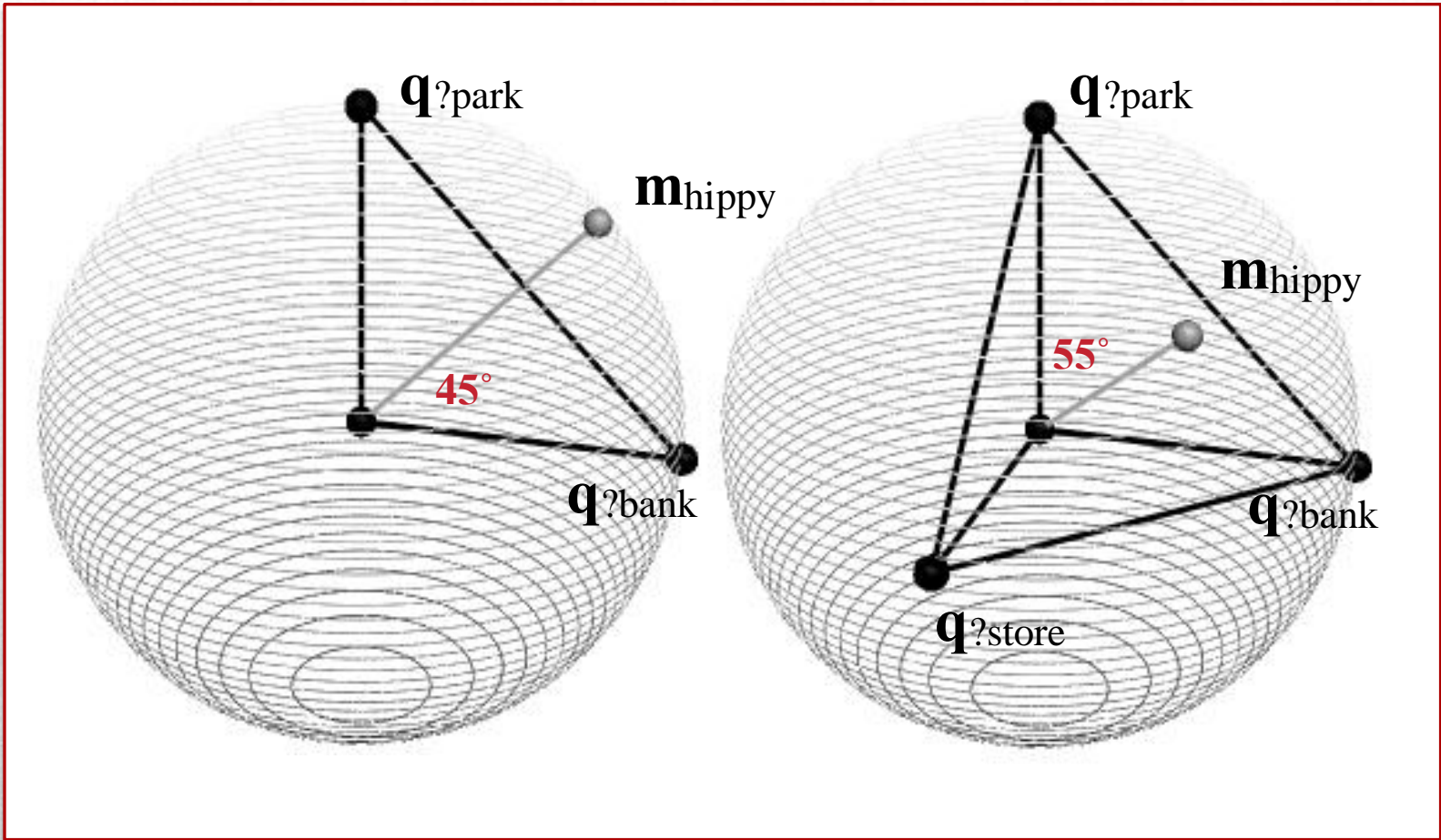
Activation & Reaction Time



HDM vs. Human



ACT-R DM vs. HDM



Fan in Vector Space

- Where f is the fan, the **cosine** between a **cue** and a **memory** vector is $f^{-1/2}$ if the vectors are perfectly orthogonal, or approximates $f^{-1/2}$ for the random vectors used by HDM.
- The cosine in HDM approximates the square-root of the probability only when the events are **equiprobable**.
- For n events with frequencies v_1 to v_n , cosine of event i is:

$$\text{cosine} = \frac{v_i}{\sqrt{v_1^2 + \dots + v_i^2 + \dots + v_n^2}}$$

Probability in Vector Space

We substitute **HDM** for **DM** in the ACT-R model of the fan effect and find that *without changing any parameters* HDM provides a good fit to the **fan effect**.

Both **DM** and **HDM** models of the fan effect claim that reaction time is a function of **conditional probability**. The vector algebra of **HDM** computes an estimate of conditional probability.

We can have HDM mimic DM's fan effect model **exactly** if we substitute **squared cosines** for S_{ji} instead of substituting A for the mean cosine.

Conclusions

HDM, by virtue of being a holographic model, has a number of capabilities for which DM is less suited:

- learning associations between concepts without having association strengths set by the modeler
- analogical or case-based reasoning
- performing tasks that require large amounts of knowledge

Conclusions

A fully vector-symbolic ACT-R would support:

- implementing partial / fuzzy matching in procedural memory
- interference in the buffers
- interfacing with outputs from a perceptual system (e.g., a deep belief network)
- being re-described at the level of neural circuitry



Future work

- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6, 451-474.
- Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, 128, 186-197.
- Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.
- Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: a distributed model of analogical mapping. *Cognitive Science*, 25, 245-286.
- Franklin, D. R. J., & Mewhort, D. J. K. (2015). Memory as a hologram: An analysis of learning and recall. *Canadian Journal of Experimental Psychology*, 69, 115-135.
- Jamieson, R. K., & Mewhort, D. J. K. (2011). Grammaticality is inferred from global similarity: A reply to Kinder (2010). *The Quarterly Journal of Experimental Psychology*, 64, 209-216. doi: 10.1080/17470218.2010.537932
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1-37.

References

- Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6, 623– 641.
- Plate, T. A. (2000). Analogy retrieval and processing with distributed vector representations. *Expert Systems: The International Journal of Knowledge Engineering and Neural Networks*, 17, 29-40. doi: 10.1111/1468-0394.00125
- Rutledge-Taylor, M. F., Pyke, A., West, R. L., & Lang, H. (2010). Modeling a three term fan effect. *Proceedings of the 10th International Conference on Cognitive Modeling*, 211-216.
- Rutledge-Taylor, M. F., Vellino, A., & West, R. L. (2008). A holographic associative memory recommender system, *Proceedings of the Third International Conference on Digital Information Management*, 87-92.
- Rutledge-Taylor, M. F., & West R. L. (2008). Modeling the fan-effect using dynamically structured holographic memory. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 385-390.
- Rutledge-Taylor, M. F., Kelly, M. A., West, R. L., & Pyke, A. A. (2014). Dynamically structured holographic memory. *Biologically Inspired Cognitive Architectures*, 9, 9-32.
- Stewart, T. C., & West, R. L. (2006). Deconstructing ACT-R. In *Proceedings of the Seventh International Conference on Cognitive Modeling* (pp. 298-303). Trieste, Italy.

References
