The memory tesseract: Mathematical equivalence between composite and separate storage memory models

Matthew A. Kelly\textsuperscript{a,*}, D.J.K. Mewhort\textsuperscript{b}, Robert L. West\textsuperscript{a}

\textsuperscript{a} Carleton University, Canada
\textsuperscript{b} Queen’s University, Canada

HIGHLIGHTS

\begin{itemize}
\item Comparative analysis points towards a unified mathematical basis for memory models.
\item MINERVA2 is proven to be equivalent to a fourth order tensor associative memory.
\item A holographic lateral inhibition network approximates MINERVA2.
\item MINERVA2 can be implemented as a fully distributed neural model.
\item MINERVA2 can be scaled up arbitrarily assuming an arbitrarily parallel computer.
\end{itemize}

ABSTRACT

Computational memory models can explain the behaviour of human memory in diverse experimental paradigms. But research has produced a profusion of competing models, and, as different models focus on different phenomena, there is no best model. However, by examining commonalities among models, we can move towards theoretical unification. Computational memory models can be grouped into composite and separate storage models. We prove that MINERVA2, a separate storage model of long-term memory, is mathematically equivalent to composite storage memory implemented as a fourth order tensor, and approximately equivalent to a fourth-order tensor compressed into a holographic vector. Building of these demonstrations, we show that MINERVA2 and related separate storage models can be implemented in neurons. Our work clarifies the relationship between composite and separate storage models of memory, and thereby moves memory models a step closer to theoretical unification.

© 2016 Elsevier Inc. All rights reserved.

Computational memory models can explain the behaviour of human memory in diverse experimental paradigms—whether it be recall or recognition, short-term or long-term retention, implicit or explicit learning. But research has produced a profusion of competing models, and, as different models focus on different phenomena, there is no best model. However, computational models of memory share many characteristics indicating wide agreement about the mathematics of how memory works. These shared characteristics can lead us towards developing a unified basis for computational models of human memory.

We argue that the class of memory models that use high dimensional vectors can be understood as belonging to a single mathematical framework. These memory models include composite vector models (e.g., Anderson, 1973; Johns, Jones, & Mewhort, 2012; Murdock, 1989), matrix models (e.g., Farrell & Lewandowsky, 2002; Humphreys, Pike, Bain, & Tehan, 1989b; Howard & Kahana, 2002; Lewandowsky & Farrell, 2008), tensor models (e.g., Humphreys, Bain, & Pike, 1989a; Osth & Dennis, 2015; Smolensky, 1990), holographic vector models (e.g., Eich, 1982; Franklin & Mewhort, 2015; Murdock, 1993), and multi-vector models such as the MINERVA2 (Hintzman, 1984) model and variants (e.g., Dougherty, Gettys, & Ogden, 1999; Jamieson, Crump, & Hannah, 2012; Jamieson & Mewhort, 2011; Kwantes, 2005; Thomas, Dougherty, Sprenger, & Harbison, 2008), the Generalized Context Model of categorization (GCM; Nosofsky, 1986, 1991), and the BEAGLE model of distributional semantics (Jones & Mewhort, 2007; see also Jones, Kintsch, & Mewhort, 2006) and variants (Kelly, Kwok, & West, 2015; Rutledge-Taylor, Kelly, West, & Pyke, 2014).

\begin{itemize}
\item Comparative analysis points towards a unified mathematical basis for memory models.
\item MINERVA2 is proven to be equivalent to a fourth order tensor associative memory.
\item A holographic lateral inhibition network approximates MINERVA2.
\item MINERVA2 can be implemented as a fully distributed neural model.
\item MINERVA2 can be scaled up arbitrarily assuming an arbitrarily parallel computer.
\end{itemize}

Keywords:
Memory
Cognitive modelling
Computational modelling
MINERVA2
Vectors
Tensors
Holographic reduced representations
HRRs
Associative memory
Neural networks
We use MINERVA 2 (Hintzman, 1984) as a starting point for developing our theoretical framework. MINERVA 2 (Hintzman, 1984) is a computational model of long-term memory (both episodic and semantic). We choose MINERVA 2 because it captures a wide variety of human memory phenomena across differing experimental paradigms and as such seems a good candidate for a basis for theoretical unification.

MINERVA 2 has been applied to a number of experimental paradigms, including judgement of frequency tasks (Hintzman, 1984), recognition tasks (Hintzman, 1984), “schema-abstraction” or category learning (Hintzman, 1984, 1986), implicit learning tasks such as artificial grammar learning (Jamieson & Mewhort, 2009, 2011), the production effect (Jamieson, Mewhort, & Hockley, 2016) as well as speech perception (Goldinger, 1998), and naming words from print (Kwantes & Mewhort, 1999).

While there are many experimental phenomena that MINERVA 2 cannot account for, our interest is not in MINERVA 2 per se, but in the broader class of models based on MINERVA 2. We refer to this broader class as MINERVA models and use MINERVA 2 to refer specifically to the model proposed by Hintzman (1984). While MINERVA 2 has many limitations, the MINERVA framework as a whole has proved a fruitful research paradigm.

Variations on MINERVA 2 address a broader range of phenomena. MINERVA-AL captures numerous associative learning phenomena from both the animal and human learning literature (Jamieson et al., 2012). Kwantes (2005) uses a MINERVA variant as a model of distributional semantics. Johns, Jamieson, Crump, Jones, and Mewhort (2016) use a MINERVA variant to model the production of natural language syntax given sentence exemplars. MINERVA-DM models judgements of likelihood to account for heuristics and biases in decision-making (Dougherty et al., 1999). The HyGene model (Thomas et al., 2008) extends MINERVA-DM to hypothesis generation and accounts for how errors in hypothesis generation lead to errors in judgement and decision-making.

In this paper, we begin by providing an introduction to MINERVA 2, followed by a comparison of the various memory models that use high-dimensional vectors. These models belong to two broad classes: composite storage and separate storage models. Composite storage models have a clear neural implementation and are invariant in scale with respect to the number of memories stored. Conversely, distributed storage models grow with the number of memories or concepts stored in the model and do not have an established neural interpretation.

To address concerns about the scalability and neural realization of separate storage models, and to move towards a unified theoretical framework for memory models, we present a proof of exact mathematical equivalence between MINERVA 2 (a separate storage model) and an auto-associative fourth-order tensor model (a composite storage model). We refer to the tensor as a “memory tesseract” as it is a matrix with four equal dimensions. We also prove that MINERVA is approximately equivalent to a holographic approximation to the memory tesseract.

To illustrate the behaviour of the memory models, we present a set of simulations on artificial data. We also compare performance of the holographic approximation to Johns et al.’s (2016) MINERVA model of a sentence production task. We find that the holographic approximation provides a means of implementing MINERVA as a memory system that is invariant in scale with respect to the number of experiences stored, but does so at the cost of increased noise from the compression of the memory traces into a smaller data structure. Also, to be feasibly implemented on very large-scale tasks, the holographic approximation needs to be simulated on a massively parallel computer, such as a neuromorphic computer.

This work clarifies the relationship between MINERVA and other memory models that use high dimensional vectors. This work serves to demonstrate that MINERVA can potentially be used as a basis for unifying high dimensional vector memory modelling, that MINERVA is scalable to arbitrarily long-term learning if implemented on a massively parallel computer, and that MINERVA can plausibly be realized in neurons.

## 1. How does MINERVA work?

In MINERVA, each individual experience, or episode, is represented by a high dimensional vector, a list of features represented by numerical values. Memory is a table where each row is a vector representing an episodic trace, a stored experience. New experiences are stored as new rows in the memory table. New experiences do not need to be novel. A repeated experience is also stored as a new row, separate from previous instances of that experience.

In MINERVA, memory retrieval is not a look-up process, it is a reconstruction process. In the words of Tulving and Watkins (1973, p. 744), a retrieval cue “combines or interacts with the stored information to create the memory of a previously experienced event”. When a retrieval cue is presented, each vector in the table “resonates” with the cue in proportion to its similarity to the cue (Hintzman, 1986).

Similarity is computed as a normalized dot-product of the cue’s vector with the stored vector. Each stored vector is activated by its cubed similarity to the cue. Information is retrieved from memory in the form of a new vector, called an echo. The echo is a weighted sum of the vectors in the table, each vector weighted by its activation. By computing activation as the cube of similarity, the contribution of the most similar vectors (or experiences) is emphasized and that of the least similar (and least relevant) is minimized. The model uses the echo to respond as appropriate for the given task. Abstract, conceptual, semantic, and categorical information reflect aggregate retrieval over many episodic traces (e.g., Goldinger, 1998; Kwantes, 2005).

Hintzman (1984, p. 96) summarizes MINERVA 2’s key assumptions:

1. only episodic traces are stored in memory,
2. repetition produces multiple traces of an item,
3. a retrieval cue contacts all traces simultaneously,
4. each trace is activated according to similarity to the cue,
5. all traces respond in parallel, retrieved information reflects their summed output.

According to Hintzman (1990), MINERVA 2 can be understood as an artificial neural network (see Fig. 1). A layer of input nodes represent the cue, a layer of output nodes represent the echo, and between the two is a hidden layer of nodes. In the hidden layer, each node corresponds to an episodic trace. It follows that MINERVA’s hidden layer is a localist network: specific nodes represent specific pieces of information.

Modellers using MINERVA are generally agnostic as to how the model is related to the brain. No one claims that for each new experience the brain grows a new neuron that is forever singly dedicated to that particular experience. But no other interpretation of how MINERVA can be implemented in neurons has been previously proposed, leaving open the question of MINERVA’s neural plausibility.

## 2. A comparison of memory models

The memory models discussed here use vector and tensor representations to simulate the processes of storage and retrieval. Tensors are a generalization of matrices. A vector is a first order tensor, a matrix is a second order tensor, a third order tensor is a “3D matrix” or a stack of matrices, and we use the term tesseract to refer to a fourth order tensor or “4D matrix”.

---

Please cite this article in press as: Kelly, M.A., et al., The memory tesseract: Mathematical equivalence between composite and separate storage memory models. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.10.006
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات