

# The Emergence of Miller's Magic Number on a Sparse Distributed Memory

Alexandre Linhares<sup>1\*</sup>, Daniel M. Chada<sup>1</sup>, Christian N. Aranha<sup>2</sup>

<sup>1</sup> Getulio Vargas Foundation/EBAPE, Rio de Janeiro, Brazil, <sup>2</sup> Cortex Intelligence/R. Assembleia 10, Rio de Janeiro, Brazil

## Abstract

Human memory is limited in the number of items held in one's mind—a limit known as “Miller's magic number”. We study the emergence of such limits as a result of the statistics of large bitvectors used to represent items in memory, given two postulates: i) the Sparse Distributed Memory; and ii) chunking through averaging. Potential implications for theoretical neuroscience are discussed.

**Citation:** Linhares A, Chada DM, Aranha CN (2011) The Emergence of Miller's Magic Number on a Sparse Distributed Memory. PLoS ONE 6(1): e15592. doi:10.1371/journal.pone.0015592

**Editor:** Colin Allen, Indiana University, United States of America

**Received:** August 16, 2010; **Accepted:** November 13, 2010; **Published:** January 5, 2011

**Copyright:** © 2011 Linhares et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work has been supported by grant E-26/110.790/2009 from the Federação de Apoio à Pesquisa do Estado do Rio de Janeiro (FAPERJ-[www.faperj.br](http://www.faperj.br)), grants 470341/2009-2 and 301207/2009-7 from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-[www.cnpq.br](http://www.cnpq.br)) of the Brazilian Ministry of Science & Technology, the Programa de Apoio à Pesquisa (PROPSQUISA) of the Getulio Vargas Foundation ([www.fgv.br](http://www.fgv.br)), and Cortex Intelligence ([www.cortex-intelligence.com](http://www.cortex-intelligence.com)), due to the employment of its CTO, Dr. C.N. Aranha. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Dr. Aranha is Chief Technology Officer of Cortex Intelligence, yet the organization had no role or intervention in the research and publication of this study, nor does it intend to pursue interests in consultancy, patents, products in development, or marketed products. Furthermore, the presence of Cortex Intelligence among the funders does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

\* E-mail: alexandre.linhares@fgv.br

## Introduction

Human short-term memory is severely limited. While the existence of such limits is undisputed, there is ample debate concerning their nature. Miller [1] described the ability to increase storage capacity by grouping items, or “chunking”. He argued that the span of attention could comprehend somewhere around seven information items. Chunk structure is recursive; as chunks may contain other chunks as items: Paragraphs built out of phrases built out of words built out of letters built out of strokes. This mechanism is used to explain the cognitive capacity to store a seemingly endless flux of incoming, pre-registered, information, while remaining unable to absorb and process new (non-registered) information in highly parallel fashion.

Miller's ‘magic number seven’ has been subject of much debate over the decades. Some cognitive scientists have modeled such limits by simply using (computer-science) “pointers”, or “slots” (e.g. [2,3]—see [4,5] for debate). However, such approaches do not seem plausible given the massively parallel nature of the brain, and we believe memory limits are an emergent property of the neural architecture of the human brain. As Hofstadter put it a quarter of a century ago [6]: the “problem with this [slot] approach is that it takes something that clearly is a very complex consequence of underlying mechanisms and simply plugs it in a complex structure, bypassing the question of what those underlying mechanisms might be.”(p. 642)

Our objective in this paper is to study these memory limits as emergent effects of underlying mechanisms. We postulate two mechanisms previously discussed in the literature. The first is a mathematical model of human memory brought forth by Kanerva [7], called Sparse Distributed Memory (SDM). We also presuppose, following [8], an underlying mechanism of chunking through

averaging. It is not within the scope of this study to argue for the validity of SDM as a cognitive model; for incursions on this broader topic, we refer readers to [9–11], which discuss the plausibility of this Vector Symbolic Architecture family of models (in which SDM is contained).

This work, while similar in its mathematical foundations, is different from previous capacity analyses: In [7], the memory capacity analysis of SDM relates to its long-term memory mechanisms, while we study its short-term memory limits. Our work also differs from that of Plate, in that, regardless of the number of items presented, the memory will only store (and subsequently retrieve) a psychologically plausible number of items. The difference becomes salient in Plate's own description [12]: “As more items and bindings are stored in a single HRR the noise on extracted items increases. If too many associations are stored, the quality will be so low that the extracted items will be easily confused with similar items or, in extreme cases, completely unrecognizable”(p. 139). Plate is focused on long-term memory; and we will focus on Miller's STM limits.

A number of theoretical observations are drawn from our computations: i) a range of plausible numbers for the dimensions of the memory, ii) a minimization of a current controversy between different ‘magic number’ estimates, and iii) potential empirical tests of the chunking through averaging assumption. We should start with a brief description of our postulates: i) the SDM, and ii) chunking through averaging.

## Sparse Distributed Memory

The Sparse Distributed Memory (SDM), developed in [7], defines a memory model in which data is stored in distributed fashion in a vast, sparsely populated, binary address space. In this model, (a number of) neurons act as *address decoders*. Consider the

space  $\{0,1\}^N$ : SDM's address space is defined allowing  $2^N$  possible locations, where  $N$  defines both the word length and the number of dimensions of the space: the memory holds binary vectors of length  $N$ . In SDM, the data is the same as the medium in which it is stored (i.e. the stored items are  $N$ -bit vectors in  $N$ -dimensional binary addresses).

SDM uses Hamming distance as a metric between any two  $N$ -bit vectors (hereafter memory items, items, elements, or bitstrings—according to context). Neurons, or *hard locations* (see below), in Kanerva's model, hold random bitstrings with equal probability of 0's and 1's—Kanerva [13,14] has been exploring a variation of this model with a very large number of dimensions (around 10000). (With the purpose of encoding concepts at many levels, the Binary Spatter Code—or BSC—, shares numerous properties with SDM.) By using the Hamming distance as a metric, one can readily see that the average distance between any two points in the space is given by the binomial distribution, and approximated by a normal curve with mean at  $N/2$  with standard deviation  $\sqrt{N}/2$ . Given the Hamming distance, and large  $N$ , most of the space lies close to the mean. A low Hamming distance between any two items means that these memory items are associated. A distance that is close to the mean  $N/2$  means that the memory items are orthogonal to each other. This reflects two facts about the organization of human memory: i) *orthogonality of random concepts*, and ii) *close paths between random concepts*.

*Orthogonality of random concepts*: the vast majority of concepts is orthogonal to all others. Consider a non-scientific survey during a cognitive science seminar, where students asked to mention ideas unrelated to the course brought up terms like *birthdays*, *boots*, *dinosaurs*, *fever*, *executive order*, *x-rays*, and so on. Not only are the items unrelated to cognitive science, the topic of the seminar, but they are also unrelated to each other.

*Close paths between concepts*: The organization of concepts seems to present a 'small world' topology—for an empirical approach on words, for instance, see [15]. For any two memory items, one can readily find a stream of thought relating two such items ("Darwin gave *dinosaurs* the *boot*"; "she ran a *fever* on her *birthday*"; "isn't it time for the Supreme Court to *x-ray* that *executive order*?" ... and so forth). Robert French presents an intriguing example in which one suddenly creates a representation linking the otherwise unrelated concepts of "coffee cups" and "old elephants" [16]. In sparse distributed memory, any two bitstrings with Hamming distance around  $N/4$  would be extremely close, given the aforementioned distribution. And  $N/4$  is the expected distance of an average point between two random bitstrings.

Of course, for large  $N$  (such as  $N \geq 100$ ), it is impossible to store all (or even most) of the space—the universe is estimated to carry a storage capacity of  $10^{90}$  bits ( $10^{120}$  bits if one considers quantum gravity) [17]. It is here that Kanerva's insights concerning sparseness and distributed storage and retrieval come into play:  $2^{20}$ —or a number around one million—physical memory locations, called hard locations, could enable the representation of a large number of different bitstrings. Items of a large space with, say,  $2^{1000}$  locations would be stored in a mere  $2^{20}$  hard locations—the memory is indeed sparse.

In this model, every single item is stored in several hard locations, and can, likewise, be retrieved in distributed fashion. Storage occurs by distributing the item in every hard location within a certain threshold 'radius' given by the Hamming distance between the item's address and the associated hard locations. Different threshold values for different numbers of dimensions are used (in his examples, Kanerva used 100, 1000 and 10000 dimensions). For  $N = 1000$ , the distance from a random point of the space to its nearest (out of the one million) hard locations will

be approximately 424 bits [7] (p.56). In this scenario, a threshold radius of 451 bits will define an *access sphere* containing around 1000 hard locations. In other words, from any point of the space, approximately 1000 hard locations lie within a 451-bit distance. All of these accessible hard locations will be used in storing and retrieving items from memory. We therefore define the function  $A : \{0,1\}^N \times \{1,2,\dots,N\} \rightarrow 2^{\{0,1\}^N}$  and a hard location  $\psi_x \in A(x,R)$  iff  $\psi_x \in \{0,1\}^N \wedge H(x,\psi_x) \leq R$ , where  $A$  defines an access radius around  $x$  of size  $R$  (451 if  $N = 1000$ ;  $H$  is the Hamming distance).

A brief example of a storage and retrieval procedure in SDM is in order: to store an item  $x$  at a given (virtual) location  $\zeta$  (in sparse memory) one must activate every hard location within the access sphere of  $\zeta$  (see below) and store the datum in each one. Hard locations carry  $N$  adders, one for each dimension. To store a bitstring  $x$  at a hard location  $\psi$ , one must iterate through the adders of  $\psi$ : If the  $i$ -th bit of  $x$  is 1, increment the  $i$ -th adder of  $\psi$ , if it is 0, decrement it. Repeating this for all hard locations in  $\zeta$ 's access sphere will distribute the information in  $x$  throughout these hard locations.

Retrieval of data in SDM is also massively collective and distributed: to peek the contents of each hard location, one computes its related bit vector from its adders, assigning the  $i$ -th bit of  $\psi$  as a 1 or 0 if the  $i$ -th adder is positive or negative, respectively (a coin is flipped if it is 0). Notice, however, that this information in itself is meaningless and may not correspond to any one specific datum previously registered. To read from a location  $x$  in the  $\{0,1\}^N$  address space, one must activate the hard locations in the access sphere of  $x$  and gather each related bit vector. The stored datum will be the majority rule decision of all activated hard locations' related bit vectors. If, for the  $i$ -th bit, the majority of all bit vectors is 1, the final read datum's  $i$ -th bit is set to 1, otherwise to 0. Thus, "SDM is distributed in that many hard locations participate in storing and retrieving each datum, and one hard location can be involved in the storage and retrieval of many data" [18] (p. 342).

All hard locations within an access radius collectively point to an address. Note also that this process is iterative. The address obtained may not have information stored on it, but it provides a new access radius to (possibly) converge to the desired original address. One particularly impressive characteristic of the model is its ability to simulate the "tip-of-tongue" phenomenon, in which one is certain about some features of the desired memory item, yet has difficulty in retrieving it (sometimes being unable to do so). If the requested address is far enough from the original item (209 bits if  $N = 1000$ ), iterations of the process will not decrease the distance—and time to convergence goes to infinity.

The model is robust against errors for at least two reasons: i) the contribution of any one hard location, in isolation, is negligible, and ii) the system can readily deal with incomplete information and still converge to a previously registered memory item. The model's sparse nature dictates that any point of the space may be used as a storage address, whether or not it corresponds to a hard location. By using about one million hard locations, the memory's distributed nature can "virtualize" the large address space. The distributed aspect of the model allows such a virtualization. Kanerva [7] also discusses the biological plausibility of the model, as the linear threshold function given by the access radius can be readily computed by neurons, and he suggests the interpretation of some particular types of neurons as address decoders. Given these preliminaries concerning the Sparse Distributed Memory, we should now proceed to our second premise: *chunking through averaging*.

### Chunking through averaging

To chunk items, the majority rule is applied to each bit: given  $v$  bitstrings to be chunked, for each of the  $N$  bits, if the majority is 1, the resulting bitstring's chunk bit is set to 1; otherwise it is 0. In case of perfect ties (no majority), a coin is flipped.

We have chosen the term ‘chunking’ to describe an averaging operation, and ‘chunk’ to describe the resulting bitstring, because, through this operation, the original components generate a new one to be written to memory. The reader should note, in SDM’s family of high-dimensional vector models, called Vector Symbolic Architectures (VSA), the operation that generates composite structures is commonly known as superposition [10–12].

Obviously, this new chunked bitstring may be closer, in terms of Hamming distance, to the original elements, than the mean distance  $N/2$  between random elements (500 bits if  $N = 1000$ ), given a relatively small  $v$ . The chunk may then be stored in the memory, and it may be used in future chunking operations, allowing, thus, for recursive behavior. With these preliminaries, we turn to numerical results in the analysis section.

### Analysis

#### Computing the Hamming distance from a chunk $\alpha$ to items

Let  $\zeta = \{\zeta_1, \zeta_2, \dots, \zeta_v\}$  be the set of bitstrings to be chunked into a new bitstring,  $\alpha$ . The first task is to find out how the Hamming distance is distributed between this averaged  $\alpha$  bitstring and the set  $\zeta = \{\zeta_1, \zeta_2, \dots, \zeta_v\}$  of bitstrings being chunked. This is, as discussed, accomplished through majority rule at each bit position. Imagine that, for each separate dimension, a supreme court will cast a decision with each judge choosing yes (1) or no (0). If there is an even number of judges, a fair coin will be flipped in the case of a tie. Given that there are  $v = |\zeta|$  votes cast, how many of these votes will fall in the minority side? (Each minority-side vote adds to the Hamming distance between an item  $\zeta_i$  and the average  $\alpha$ .)

Note that the minimum possible number of *minority votes* is one, and that it may occur with either 3 votes cast or two votes and a coin flip. If there are two minority votes, they may stem from either 5 votes or 4 votes and a coin flip, and so forth. We thus have that, for  $v$  votes, the maximum minority number is given by  $\lfloor v/2 \rfloor$  (and the ambiguities between an odd number of votes versus an even number of votes plus a coin flip are resolved by considering  $2\lfloor v/2 \rfloor + 1$  total votes). This leads to independent Bernoulli trials, with success factor  $p = 1/2$ , and the constraint that the minority view differs from the majority bit vote. Let  $X$  be a random variable with the number of minority votes. Obviously in this case,  $P(1 \leq X \leq \lfloor v/2 \rfloor) = P(X \leq \lfloor v/2 \rfloor - 1)$ , hence we have, for  $v$  items, the following cumulative distribution function of minority votes [19]:

$$P(X \leq \lfloor v/2 \rfloor - 1) = \sum_{i=0}^{\lfloor v/2 \rfloor - 1} \binom{2\lfloor v/2 \rfloor}{i} p^i (1-p)^{2\lfloor v/2 \rfloor - i}$$

$$= \sum_{i=0}^{\lfloor v/2 \rfloor - 1} \frac{(2\lfloor v/2 \rfloor)!}{i!(2\lfloor v/2 \rfloor - i)!} \frac{1}{2} = 4^{-\lfloor v/2 \rfloor} \sum_{i=0}^{\lfloor v/2 \rfloor - 1} \frac{(2\lfloor v/2 \rfloor)!}{i!(2\lfloor v/2 \rfloor - i)!}$$

While we can now, given  $v$  votes, compute the distribution of minority votes, the objective is not to understand the behavior of these minority bits *in isolation*, i.e., per dimension on the chunking process. We want to compute the number of dimensions to (in a

psychologically and neurologically plausible way) store and retrieve around  $M$  items—Miller’s number of retrievable elements—through an averaging operation. Hence we need to compute the following:

- (i) Given a number of dimensions  $N$  and a set  $\zeta$  of items, the probability density function of the Hamming distance from  $\alpha$  to the chunked elements  $\zeta_i$ ,
- (ii) A threshold  $T$ : a number of dimensions in which, if an element  $\zeta_i$ 's Hamming distance to  $\alpha$  is farther from that point, then  $\zeta_i$  cannot be retrieved,
- (iii) As  $|\zeta|$  grows, how many elements remain retrievable?

Given bitstrings with dimension  $N$ , suppose  $v = |\zeta|$  elements have been chunked, generating a new bitstring  $\alpha$ . Let  $H_N(\alpha, \zeta_i)$  be the Hamming distance from the chunked element  $\alpha$  to  $\zeta_i$ , the  $i$ -th element of  $\zeta$ . What is the distance from  $\alpha$  to elements in  $\zeta$ ? Here we are led to  $N$  Bernoulli trials with success factor  $p_{|\zeta|}$ . Since  $N$  is large,  $H_N(\alpha, \zeta_i)$  for  $i = \{1, 2, \dots, v\}$  can be approximated by a Normal distribution, we may use  $\mu = Np_{|\zeta|}$  and  $\sigma = \sqrt{Np_{|\zeta|}(1-p_{|\zeta|})}$ . To model human short term memory’s limitations, we want to compute a cutoff threshold  $T_N$  which will guarantee retrieval of around  $M$  items averaged in  $\alpha$  and “forget” an item  $\zeta_i$  if  $H_N(\alpha, \zeta_i) > T_N$ —where  $M$  is Miller’s limiting number. Hence to guarantee retrieval of around 95% ( $2\sigma$ ) of  $M$  items, we have  $T_{N,2\sigma} = Np_M + 2\sqrt{Np_M(1-p_M)}$ , where  $p_M$  is the success factor corresponding to  $M$ . Note that Cowan [20] has argued for a “magic number” estimate of  $4 \pm 1$  items—and the exact cognitive limit is still a matter of debate. The success factor for 4 (or 5) elements is  $p_{\{4,5\}} = .3125$ ; and for 6 (or 7) elements it is  $p_{\{6,7\}} = .34375$ . By fixing the success factor at plausible values of  $M$  (at  $\{4,5\}$ , or at an intermediary value between  $\{4,5\}$  and  $\{6,7\}$ , or at  $\{6,7\}$ ), different threshold values  $T_{N,2\sigma}$  are obtained for varying  $N$ , as shown in Table 1. In the remainder of this study, we use the intermediary success factor  $p_M = .328125 = 21/64$  for our computations; again without loss of generality between different estimates of  $M$ .

We thus have a number of plausible thresholds and dimensions. We can now proceed to compute the plausibility range: Despite the implicit suggestion in Table 1 that any number of dimensions might be plausible, how does the behavior of these  $(N, T_{N,2\sigma})$  combinations vary as a function of the number of presented elements,  $|\zeta|$ ?

#### Varying the number of presented items

Consider the case of information overload, when one is presented with a large set of items. Suppose one were faced with dozens, or hundreds, of distinct items. It is not psychologically plausible that a large number of elements should be retrievable. For an item  $\zeta_i$  to be impossible to retrieve, the distance between the averaged item  $\alpha$  and  $\zeta_i$  must be higher than the threshold point of the corresponding  $N$ . When we have an increasingly large set of presented items, there will be information loss in the chunking mechanism, but it should still be possible to retrieve some elements within plausible psychological bounds.

Figure 1(a) shows the behavior of three representative sizes of  $N$ : 100, 212 and 1000 dimensions. (100 and 1000 were chosen because these are described in Kanerva’s original examples of SDM.)  $N = 212$  has shown to be the most plausible number of dimensions, preserving a psychologically plausible number of items after presentations of different set sizes. It is clear that  $N = 100$  quickly diverges, retaining a high number of items in a chunk (as the number of presented items grows). Conversely, if  $N = 1000$ , the number of preserved memory items rapidly drops to zero, and

**Table 1.** Threshold values.

$N$	$T_{N,2\sigma}$ ( $M=4$ or $5$ )	$T_{N,2\sigma}$ (intermediary value)	$T_{N,2\sigma}$ ( $M=6$ or $7$ )
64	27.42	28.51	29.6
128	50.49	52.62	54.75
192	72.85	76.01	79.16
256	94.83	99.02	103.2
320	116.58	121.8	126.99
384	138.17	144.4	150.61
448	159.62	166.88	174.11
512	180.98	189.25	197.49
576	202.25	211.54	220.8
640	223.45	233.76	244.03
704	244.6	255.92	267.2
768	265.69	278.02	290.32
832	286.74	300.09	313.4
896	307.75	322.11	336.43
960	328.72	344.1	359.43
1024	349.66	366.05	382.4
100	40.52	42.2	43.87
1000	341.82	357.82	373.79
10000	3217.7	3375.16	3532.49

Thresholds  $T_{N,2\sigma}$  given plausible success factors and dimension combinations. doi:10.1371/journal.pone.0015592.t001

the postulated mechanisms are unable to retrieve any items at all—a psychologically implausible development. Figure 1(b) zooms in to illustrate behavior over a narrower range of  $N$ -values and a wider range of presented items. Varying the number of presented items and computing the number of preserved items (for a number of representative dimensions) yields informative results. Based on our premises, experiments show that to appropriately reflect the storage capacity limits exhibited by humans, certain ranges of  $N$  must be discarded. With too small a number of dimensions, the model will retrieve too many items in a chunk. With too large a number of dimensions, the model will retrieve at most one or two—perhaps no items at all. This is because of the higher number of standard deviations involved in the dimension sizes: for  $N=100$ , the whole space has 20 standard deviations, and  $T_{N=100,2\sigma}=42.2$  is less than 2 standard deviations below the mean—which explains why an ever growing number of items is “retrieved” (i.e., high probability of false positives). For  $N=1000$ , the space has over 63 standard deviations, and  $T_{N=1000,2\sigma}=357.82$ , is around 8.99 standard deviations below the mean. There is such a minute part of the space below  $T_{N=1000,2\sigma}$  that item retrieval is virtually impossible.

With an intermediary success factor  $p_M$  between  $p_4$  and  $p_7$  established by the cognitive limits 4 and 7, we have computed the number of dimensions of a SDM as lying in the vicinity of 212 dimensions. Variance is minimized when  $N=212$ —and retrieval results hold psychologically plausible ranges even when hundreds of items are presented (i.e., the SDM would be able to retrieve from a chunk no more than nine items and at least one or two, regardless of how many items are presented simultaneously). Finally, given that this work rests upon the chunking through averaging postulate, in the next section we will argue that the postulated mechanism is not only plausible, but also empirically testable.

## Results and Discussion

### The chunking through averaging postulate

Consider the assumption of chunking through averaging. We propose that it is plausible and worthy of further investigation, for three reasons.

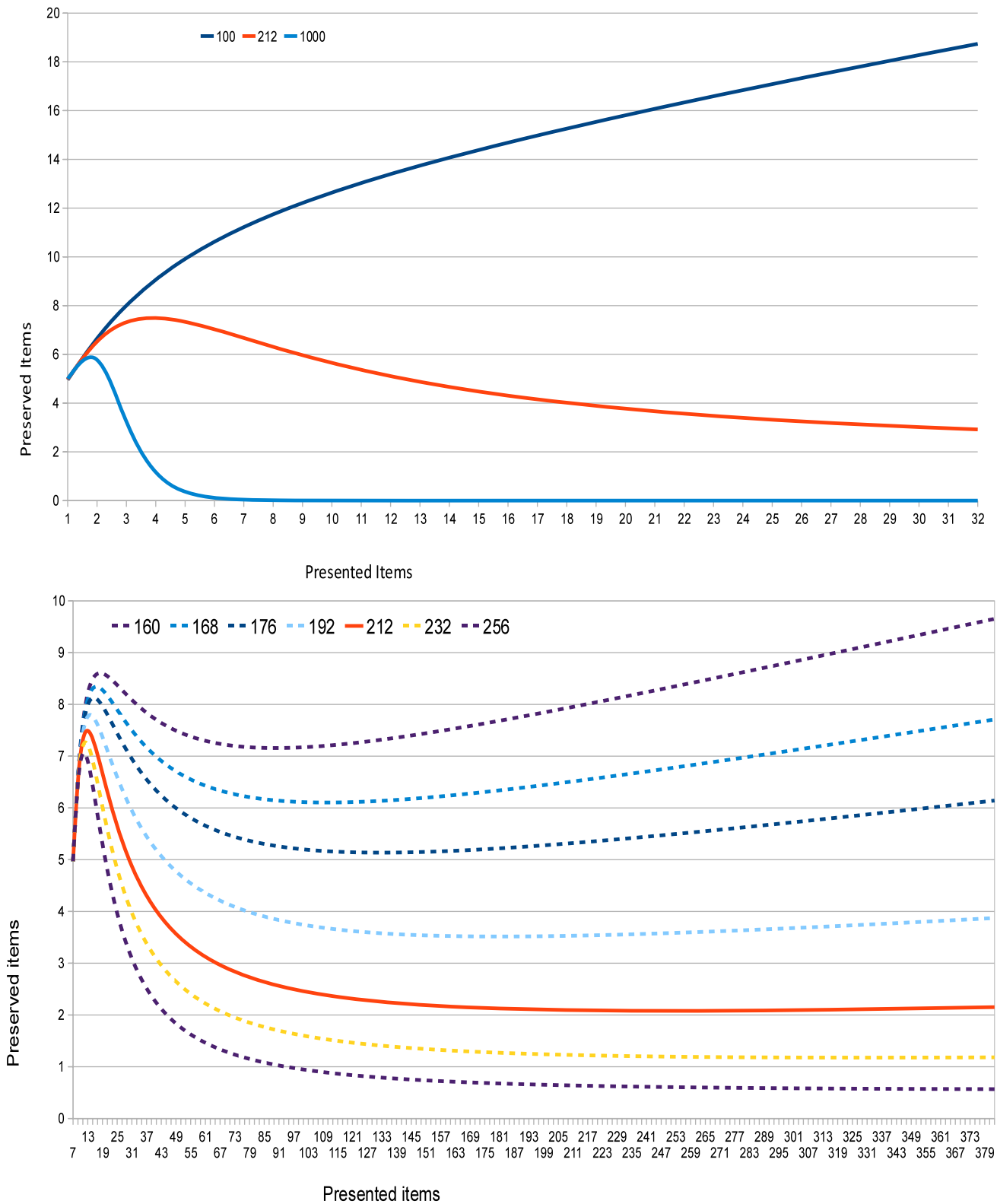
First, it minimizes the current controversy between Miller’s estimations and Cowan’s. The disparity between Miller’s  $7 \pm 2$  or Cowan’s  $4 \pm 1$  observed limits may be a smaller delta than what is argued by Cowan. Our “chunking-through-averaging” premise may provide a simpler, and perhaps unifying, position to this debate. If chunking 4 items has the same probability as 5 items, and chunking 6 items is equivalent to chunking 7 items, one may find that the ‘magic number’ constitutes one cumulative probability degree (say, 4-or-5 items) plus or minus one (6-or-7 items).

A mainstream interpretation of the above phenomenon may be that, as with any model, SDM is a simplification; an idealized approximation of a presumed reality. Thus, one may see it as insufficiently complete to accurately replicate the details of true biological function due to, among other phenomena, inherent noise and spiking neural activity. In this case, one would interpret it as a weakness, or an inaccuracy inherent to the model. An alternative view, however improbable, may be that the model is accurate in this particular aspect, in which case, the assumption minimizes the current controversy between Miller’s estimations and Cowan’s.

The success factors computed above show that for either 4 or 5 items, we have  $p=.3125$ , while for 6 or 7 items we have  $p=.34375$ . If we assume an intermediary value of  $p$ —which is reasonable, due to noise or lack of synchronicity in neural processing—the controversy vanishes. We chose to base our experiments on the mean value ( $p=21/64=.328125$ ), and the results herein may be adapted to other estimates as additional experiments settle the debate.

Moreover, a chunk  $\alpha$  tends to be closer to the  $\zeta$ ; chunked items than these items are between themselves. For example, with  $|\zeta|=5$  and  $N=212$ , the Hamming distance between a chunk and a random item is drawn from a distribution with  $\mu=N21/64$  and  $\sigma=(\sqrt{N}\sqrt{903})/64$ ; in here, from the point of view of the chunked item  $\alpha$ , the closest 1% of the space lies at 53 bits, while 99% of the space lies at 84 bits. Contrast this with the distances between any two random, orthogonal, items, which are drawn from  $\mu=N/2$  and  $\sigma=\sqrt{N}/2$ : from the point of view of a random item, the closest 1% of the space lies at 89 bits, while 99% of the space lies at 122. This disparity reflects the principles of *orthogonality between random concepts* and of *close paths between concepts* (or small worlds [15]): the distance between 2 items from any 5 is large, but the distance to the average of the set is small. Of course, as  $|\zeta|$  grows, the distance to  $\alpha$  also grows (since  $\lim_{|\zeta| \rightarrow \infty} p_{|\zeta|}=1/2$ ), and items become irretrievable. One thing is clear: with 5 chunked items, the chance of retrieving a false positive is minute.

Finally, the assumption of chunking through averaging is empirically testable. Psychological experiments concerning the difference in ability to retain items could test this postulate. The assumption predicts that (4, 5) items, or more generally that  $(2v, 2v+1)$  for integer  $v>0$  will be registered with equal probability. It also predicts how the probability of  $2v+2$  retained items should drop in relation to  $2v+1$  if  $v>0$ . This is counterintuitive and can be measured experimentally. Note, however, two qualifications: first, as chunks are hierarchically organized, these effects may be hard to perceive in experimental settings. One would have to devise an experimental setting with



**Figure 1. Preserved items as a function of  $v$ ; selected values of  $N$ .**  
 doi:10.1371/journal.pone.0015592.g001

assurances that only chunks from the same level are retrievable—neither combinations of such chunks, nor combinations of their constituting parts. The final qualification is that, as  $v$  grows, the

aforementioned probability difference tends to zero. Because of the conjunction of these qualifications, this effect would be hard to perceive on normal human behavior.

## Concluding remarks

Numerous cognitive scientists model the limits of human short-term memory through explicit “pointers” or “slots”. In this paper we have considered the consequences of a short-term memory limit given the mechanisms of i) Kanerva’s Sparse Distributed Memory, and ii) chunking through averaging. Given an appropriate choice for the number of dimensions of the binary space, we are able to model chunks that limit active memory’s storage capacity, while allowing the theoretically endless recursive association of pre-registered memory items at different levels of abstraction (i.e., chunks may be chunked with other chunks or items, indiscriminately [1,21]). This has been pointed out in [22], however, in here we use the short-term memory limitations as a bounding factor to compute plausible ranges for  $N$ .

Some observations are noteworthy. First, our work provides plausible bounds on the number of dimensions of a SDM—we make no claims concerning Kanerva’s recent work (e.g., [14]). Given our postulates, it seems that 100 dimensions is too low a number, and 1000 dimensions too high. In our computations, assuming  $p_M = 21/64$ , variance of the number of items retained (as a function of the number of presented items and at least one retrievable item) was minimized at 212 dimensions. This value was chosen as our optimal point of focus for it provided stable, psychologically plausible behavior for a wide range of set sizes. We have concentrated on the SDM and chunking through averaging postulates, yet future research could also look at alternative neural models; for it is certain that the brain does not use explicit slots or pointers when items are chunked. One can reasonably argue: what good can come from replacing one magic number with another? There are two potential benefits: first, by fixing parameter  $N$ , we can restrict the design space of SDM simulations and ensure that a psychologically plausible number of items is chunked. Another advantage is theoretical: the number 212 suggests that we should look for neurons that seem to have, or respond majoritarily to, such a number of active inputs in their linear threshold function.

Of course, a single 212 bit vector in SDM does not encode meaningful content at all. The existence of a bitstring can only be

meaningful in relation to other bitstrings close to it. Consider, for instance, an A4 sheet of paper, of size 210mm×297mm (8.3in×11.7in). A 1200×1200 dots-per-inch printer holds less than  $2^{28}$  potential dots in an entire sheet. While the space of possible black and white printed A4 sheets is a very large set of  $2^{139838400}$  possible pages, the vast majority of them, rather like the library of Babel, are composed of utter gibberish. Any single dot needs only 28 bits to be described, and because the dots usually cluster into strokes, chunks can be formed. Moreover, because strokes cluster to form fonts, which cluster to form words, which cluster to form phrases and paragraphs; combinations of large sets of 212 dimensional bitstrings can encode the meaningful content of pages and books—provided those items have been previously chunked in the reader’s mind. Without chunks there can be no meaning; this paragraph, translated to Yanomami (assuming that’s possible), would become unreadable to its intended audience and to its authors.

Sparse Distributed Memory holds a number of biologically and psychologically plausible characteristics. It is associative, allowing for accurate retrieval given vague or incomplete information (which is relevant given the potential for asynchronous behavior [23]); it is readily computable by neurons; it seems suitable for storage and retrieval of low-level sensorimotor information [24], it is a plausible model of the space of human concepts, and it exhibits a phenomenon strikingly similar to the tip-of-the-tongue situation. With the results presented herein, sparse distributed memory also reflects the natural limits of human short-term memory.

## Acknowledgments

The authors would like to thank Eric Nichols for numerous valuable comments.

## Author Contributions

Conceived and designed the experiments: AL DMC CNA. Performed the experiments: AL DMC CNA. Analyzed the data: AL DMC. Wrote the paper: AL DMC CNA.

## References

- Miller GA (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol Rev* 63: 81–97.
- Gobet F, Simon HA (2000) Five seconds or sixty? Presentation time in expert memory. *Cognitive Sci* 24: 651–682.
- Gobet F, Lane PCR, Croker S, Cheng PCH, Jones G, Oliver I, Pine JM (2001) Chunking mechanisms in human learning. *Trends Cognitive Sci* 5: 236–243.
- Linhares A, Brum P (2007) Understanding our understanding of strategic scenarios: What role do chunks play? *Cognitive Sci* 31: 989–1007.
- Linhares A, Freitas AETA (2010) Questioning Chase and Simon’s (1973) “Perception in chess”: The “experience recognition” hypothesis. *New Ideas Psychol* 28: 64–78.
- Hofstadter DR (1985) *Metamagical Themas, Questing for the Essence of Mind and Pattern*, Basic Books, New York, 1985.
- Kanerva P (1988) *Sparse Distributed Memory*, MIT Press, Cambridge.
- Kanerva P (1998) Large patterns make great symbols: An example of learning from example. In: *Workshop proceedings of NIPS98: Neural Information Processing Systems Conference*, 4–5, Denver, Colorado, USA.
- Kanerva P (1993) *Sparse Distributed Memory and related models*. In MHAssoun, ed. *Associative Neural Memories: Theory and Implementation*. New York: Oxford University Press. pp 50–76.
- Stewart TC, Eliasmith M (2009) *Compositionality and biologically plausible models* Oxford Handbook of Compositionality.
- Gayler RW (2003) Vector Symbolic Architectures answer Jackendoff’s challenges for cognitive neuroscience. In PeterSlezak, ed. *ICCS/ASCS International Conference on Cognitive Science*. Sydney, Australia: University of New South Wales. pp 133–138.
- Plate TA (2003) *Holographic reduced representations, Distributed representations for cognitive structures*, CSLI Publications, Stanford, CA.
- Kanerva P (1994) *The Spatter Code for encoding concepts at many levels*. In MMarinaro, PGMorasso, 1 vol, eds. *ICANN ’94, Proceedings of International Conference on Artificial Neural Networks*, vol. 1. London: Springer-Verlag. pp 226–229.
- Kanerva P (2009) Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Comput* 1: 139–159.
- Cancho RF, Sole R (2001) The small world of human language. *Proc Royal Soc B* 268: 2261–2265.
- French RM (2000) When Coffee Cups Are Like Old Elephants, or Why Representation Modules Don’t Make Sense, in *Understanding Representation in the Cognitive Sciences 2000*, Part 3. pp 93–100. DOI: 10.1007/978-0-585-29605-0\_10.
- Lloyd S (2002) Computational capacity of the Universe. *Phys Rev Lett* 88: 237901–1:4.
- Anwar A, Franklin S (2003) Sparse distributed memory for “conscious” software agents. *Cognitive Sys Res* 4: 339–354.
- Boland PJ (1989) Majority Systems and the Condorcet Jury Theorem. *The Statistician* 38: 181–189.
- Cowan N (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav Brain Sci* 24: 87–185.
- Linhares A (2000) A glimpse at the metaphysics of Bongard problems. *Artif Intell* 121: 251–270.
- Kanerva P (1996) Binary spatter-coding of ordered k-tuples. in *ICANN96, Artificial Neural Networks* Springer. pp 896–873.
- Borisyuk R, Denham M, Hoppensteadt F, Kazanovich Y, Vinogradova O (2000) An oscillatory neural network model of sparse distributed memory and novelty detection. *Biosystems* 58: 265–272.
- Assisi C, Stopfer M, Laurent G, Bazhenov M (2007) Adaptive regulation of sparseness by feedforward inhibition. *Nature Neurosci* 10: 1176–1184.