LES CAHIERS DU LANCI

CAN HOLOGRAPHIC REDUCED REPRESENTATIONS OPERATIONS BE USED WITH MODAL REPRESENTATIONS? AN EXPERIMENT.

Jean-Frédéric de Pasquale
Pierre Poirier

UQÀM
Université du Québec à Montréal

No 2014-01

Le Laboratoire d'ANalyse Cognitive de l'Information (LANCI) effectue des recherches sur le traitement cognitif de l'information. La recherche fondamentale porte sur les multiples conceptions de l'information. Elle s'intéresse plus particulièrement aux modèles cognitifs de la classification et de la catégorisation, tant dans une perspective symbolique que connexionniste.

La recherche appliquée explore les technologies informatiques qui manipulent l'information. Le territoire privilégié est celui du texte.

La recherche est de nature interdisciplinaire. Elle en appelle à la philosophie, à l'informatique, à la linguistique et à la psychologie.

Volume 9, Numéro 2014-01 – Septembre 2014
Publication du Laboratoire d'ANalyse Cognitive de l'Information
Directeurs : Luc Faucher, Jean-Guy Meunier, Serge Robert et Pierre Poirier
Université du Québec à Montréal
Document disponible en ligne à l'adresse suivante : www.lanci.uqam.ca
Tirage : 5 exemplaires

Mise en page : Florian Ferrand

CAN HOLOGRAPHIC REDUCED REPRESENTATIONS OPERATIONS BE USED WITH MODAL REPRESENTATIONS?[1] *AN EXPERIMENT.*

Jean-Frédéric de Pasquale
Pierre Poirier
Université du Québec à Montréal

Abstract

Some cognitive theories (e.g., Thagard and Stewart 2011, Eliasmith 2013) claim that various aspects of cognition can be explained by applying the operations underlying HRRs to modal representations. Doubts based on theoretical considerations have however been raised regarding the compatibility of HRR operations and modal representations (Fisher et al. 1987). This paper aims to provide a qualitative test for the claim that HRR operations can be used with modal representations. Our results are mostly positive, while not entirely settling the issue. All types of representations (whether randomly or modally generated) processed with HRR operations fare better in high dimensions as opposed to lower dimensions. However, while the performance of modal representations remains acceptable in high dimensions, there is a distinct degradation of performance with this type of representations. Moreover, the standardization process in such cases plays a greater role than it should considering its low neural plausibility.

1. Introduction

Holographic reduced representations (HRRs, Plate 2003) are a powerful framework that can be used to implement structured representations that are distributed and easily processed by neural networks. Descended from binding schemes for representations such as Hinton's conjunctive coding (Hinton 1990) and Smolensky's tensor products (Smolensky 1990), their main operations are the circular convolution of two vectors[2] (**a*b**), the circular correlation of

[1] We wish to thank Chris Eliasmith and Terry Stewart for answering our questions about SPA and the LANCI (UQÀM) lecture group on *How to build a brain* for a series of stimulating conversations about SPA. We take full responsibility for any error or miscomprehension in the present article.
[2] The i-th component (i=0,1,2... N-1) of the convolution **c** of **a** and **b** is found by setting $c_i=\Sigma_{j=0}^{N-1} a_j b_{(i-j) \bmod N}$.

two vectors ($\mathbf{a}\#\mathbf{b}$), the approximate inverse of a vector ($\mathbf{a}'$) and the superposition of two vectors ($\mathbf{a}+\mathbf{b}$). The algebraic properties of these operations are as follow: $\mathbf{a}*\mathbf{b}'\approx\mathbf{a}\#\mathbf{b}$, $\mathbf{a}*\mathbf{a}'\approx\mathbf{I}$ (i.e., the identity vector <1, 0, 0...>), $(\mathbf{a}*\mathbf{b})*\mathbf{c}=\mathbf{a}*(\mathbf{b}*\mathbf{c})$ (associativity), $\mathbf{a}*\mathbf{b}=\mathbf{b}*\mathbf{a}$ (commutativity), $\mathbf{a}*(\mathbf{b}+\mathbf{c})=\mathbf{a}*\mathbf{b}+\mathbf{a}*\mathbf{c}$ (distributivity of convolution over superposition). The relation $\mathbf{a}*\mathbf{b}*\mathbf{b}'\approx\mathbf{a}$ underlies much of the processing done with HRRs. With convolution, roles can be bound to concepts, as in **romeo*****agent**, which can then be superposed into proposition-like representations, as in **romeo*****agent**+**love*****verb**+**juliet*****patient**. Because basic concepts are generally generated as Gaussian vectors that have a zero dot product with each other, the convolution of the above proposition-like representation with the inverse of the HRR **agent**, that is

$$(\mathbf{romeo}*\mathbf{agent}+\mathbf{love}*\mathbf{verb}+\mathbf{juliet}*\mathbf{patient})*\mathbf{agent}'$$
$$= \mathbf{romeo}*\mathbf{agent}*\mathbf{agent}'+\mathbf{love}*\mathbf{verb}*\mathbf{agent}'+\mathbf{juiliet}*\mathbf{patient}*\mathbf{agent}'$$

will have maximal dot product with **romeo**, since $\mathbf{agent}*\mathbf{agent}'\approx\mathbf{I}$; the other terms of the sum being considered noise that can, in general, be cleaned up by a « clean-up memory » (Stewart, Tang and Eliasmith 2011). Accordingly, the previous operation can be interpreted cognitively as answering "Romeo" to the question "Who's the agent in the proposition-like representation 'Romeo loves Juliet'?"

Early examinations of convolutional memories applied to images were skeptical (Fisher et al. 1987) and lead to the verdict (by Plate himself in his 2003 book) that if HRRs are to be used with the output of perceptual systems, there should be, in order to insure that the HRRs always respect the mathematical properties necessary for them to work, a mapping between the two sets of representations (i.e. a mapping between the set of representational outputs of perceptual systems and a set of vectors with components generated as independent and identically distributed Gaussian variables, hereafter "Gaussian generated vectors" or - since the context makes the distribution obvious - just "random vectors"). In particular, the components of basic HRRs should not be correlated with each other in any way (standard correlation or cross correlations). Similarity is allowed for complex HRRs, but even then, a large number of similar items in the "lexicon" of recognized and cleaned-up items can lead to a dramatic deterioration of performance, even when dimensionality is taken into account (Plate 2003, Appendix D). An important fact about HRRs is that they function better in high dimensions, for instance in 512 or more dimensions – and this is also true for handling similarity in the lexicon.

Recently however, the claim that HRR operations can be used with modal representations appeared in a number of publication, originating mainly from the University of Waterloo. Thagard and Stewart (2011) claim that the convolution of modal representations underlies creative combinations. More radically, Eliasmith's CNRG team (Eliasmith et al. 2012, Eliasmith 2013) has developed an architecture (the Semantic Pointer Architecture, SPA) based on the notion of a *semantic pointer*: typically, a representation that can be processed through HRR operations, but that is generated from raw sensory input (at least in the case of basic semantic pointers). Not all semantic pointers are generated in this fashion, though, and in Spaun, an artificial brain that contains 2.5 millions of neurons and which is meant to showcase the modelling power of the SPA architecture, modal pointers are mapped to randomly generated pointers (or compositions of such pointers) in order to be the object of the kind of cognitive processing necessary to model high-level cognitive processes. This arrangement, which is exactly the one that Plate proposed, leaves the main question unanswered: are truly modal representations, not just random representations arbitrarily associated with such modal representations, capable of playing the symbolic role that the semantic pointer architecture promises they can?

This question is rendered more pressing by Eliasmith's claim that, by combining modal representations with HRR operations (or some other kind of vector-symbolic framework), the SPA solves the grounding problem. In the SPA, modal pointers have a shallow semantic that constitutes their modality; but they also have a deep semantic, constituted by high-dimensional, low-level representations that only these modal pointers can regenerate fully. Eliasmith writes: "I would argue that capturing deep semantics and relating them to high-level representations solves the symbol grounding problem - if we can show how those high-level representations can function like symbols." (Eliasmith, 2013, p.97)

We are aware of at least one model that empirically tests, albeit indirectly, this issue of using modal representations with HRR operations (Hunsberger et al. 2013). This model is intended to show that semantic pointers can explain human categorization, and does so with a memory trace that is the sum of the convolutions of modal, perceptual pointers generated from raw images by an autoencoder and their amodal training labels. To categorize, the system must convolve the inverse of a newly generated perceptual pointer for a new image with the memory trace, and take the label that is more similar with the result.

At first glance, Hunsberger et al.'s model seems to strongly support the view that HRR operations can indeed be used with modal representations. But we suspected that the fact that the convolved pairs in this model comprise both modal (perceptual) and amodal (label) pointers may have facilitated the performance of the model. Moreover, we considered that, whether this would turn out to be the case or not, replication and independent testing of hypotheses, here as everywhere in science, are important and so we felt the need to submit the hypothesis to our own test.

Hypotheses

Our Hypothesis H1 is that, compared to using randomly generated representations, using modal representations degrades the accuracy of HRR processing. More strongly, our hypothesis H2 is that using modal representations will result in catastrophic degradation. Truth of the stronger hypothesis implies that truly semantic pointers, i.e., structures that both are modal and can be processed quasi-symbolically, do not exist.

Operationalization

It is difficult to define modality without making amodal representations trivially impossible (for instance, if a modal representation is one that is "causally connected" with the perceptual and motor apparatus, then there doesn't seem to be any room left for neurons supporting amodal representations, and evolutionary considerations seem to rule them out – the first part of this argument has been made by Markman and Stillwell (2004) in a discussion of the amodal/multimodal distinction in Prinz's neo-empiricism). Our operationalization gets around this problem by proceeding by exemplars instead of necessary and sufficient conditions. We will consider uncontroversial the idea that the activations of the hidden layer units of a network trained on a task for which the inputs are sensory (i.e., barely processed images, sounds, textures, etc., encoded in a way that mimics a simple detector encoding – e.g., topographically organized arrays of elements specified as on or off) will qualify as modal representations. We also would argue that, even prior to learning, such hidden-units activations are modal. This is because, as was argued by Plate (2003) in his analysis of Elman's grammar-learning network, the structure of the input is often revealed by the clustering of the hidden-units activation patterns, even when no learning occurs. So there are reasons to think that even without learning, distributed representations in the hidden units of a network have something to tell us about modal representations, and particularly about the properties of modal representations that are relevant to their use with HRR operations (see the "Motivation for the kinds of representations chosen" section below for more on this subject.)

To test H2, we will say that, provided our results are representative of the relevant class of similar experiments, semantic pointers are impossible (even with chaining schemes that have been proposed in Eliasmith 2013) if accuracy drops to below 50%[3]. We do not read too much into this number – it is simply a low enough number for there to be no doubt that processing is seriously impaired if the accuracy falls below it. To reject H2, the performance of modal representations must be above 50% on the 8-pairs task (see below) for at least one of the numbers of dimensions tested. We motivate the number 8 by the fact that Eliasmith claims that accuracy is 99% for 1 to 8 pairs, which plays a role in the theory of chained decoding and also because, as Eliasmith writes, "Miller's (1956) classic limits of seven plus or minus two on working memory fall in about the same range" (Eliasmith 2013, p.143).

Method[4]

Generation of representations

To test our hypotheses, we needed something that would have the qualities of modal representations. We used Rueckl et al.'s (1989) task, in which a two-layers MLP network has a 5x5 retina and must output the identity (the "what" task) of the object projected on the retina (9 possibilities, represented locally) as well as its location (the "where" task) (9 possibilities, as the object is contained in a 3x3 submatrix). Rueckl et al.'s goal was to demonstrate the usefulness of modularity for learning, but that is not our goal here: we only use their task as a source of neurally processed modal representations. More specifically, for each experiment involving learned representations, we train two networks, one on the *what* task, one on the *where* task[5], we record the 81 patterns of hidden units for each network, and we then store them as 2x81=162 basic representations. These representations must be preprocessed to be suitable for HRR processing. Do to so, we subtract their average and normalize them, so that components will

[3] Random guessing would lead to an accuracy of about 0.6% for all tasks.
[4] The generation and processing of the representations, as well as basic result data aggregation, is done by an in-house software in Java; we compute the convolution by using the definition (i.e., with real numbers) instead of using the direct and inverse Fast Fourier Transforms as Eliasmith proposes in (Eliasmith 2013).
[5] The settings are as follows: each network sees the 81 patterns 500 times, the learning rate is set to 0.05, the momentum is set to 0.

have a sample average of zero and a variance of approximately 1/N. This can be done in two different ways: by standardizing the set of patterns as a whole, or by standardizing each subset separately; we will see that this makes a difference. For comparison purposes, we generated three other sets of 2x81 basic representations: one, which we call "random-modal," where the same procedure is applied but no training occurs (the relevance of this kind of representations was motivated in the "Operationalization" section), another, which we call "mixed," where half the patterns are modal and half are random, and a third where all the patterns are random, and which we thus call "random".

Motivation of the kind of representations chosen

Learned representations and random-modal representations have properties that can challenge traditional HRR conditions of functioning. First, their components are not independent of one another. Second, different representations may be similar to each other. This is clear for the *where* task, where similarly located objects will cause similar activation in the hidden layer, whether there is any learning or not. But learning also introduces similarity (up to a point) for the *what* task. Even without learning, it is possible, in networks in which processing preserves the topological structure of the input, that cross-correlations between representations of objects at different positions (those being simply shifted versions of one another) will be picked up by convolution of the two representations – which can precisely be used to detect such shifts in a signal (this, however, is highly unlikely in our networks, since scrambling the order of the hidden layer neurons while preserving their weights gives a functionally equivalent network).

Motivation of the preprocessing

Some might have concerns about the fact that we standardize the representations. We must remind the reader that proponents of the semantic pointers idea operate within the Neural Engineering Framework, a framework that is useful for engineering spiking neural networks. In that framework, what a population of neurons represents is determined via an encoding function and a decoding function. But if **x** can be decoded from the population, then so can f**(x)** (provided there is enough heterogeneity in the population[6]; Eliasmith and Anderson (2003, p. 8-

[6] In general the decoding may not be as good, but in the present situation, where all we have to do is apply a very simple linear transformation and add a constant term, we have good reason to think it will be. While this seems

9) recognize this indetermination). Suppose there is a set of pointers **x** that have a certain distribution (derived from the distribution of sensory stimuli) so that the mean of the components x_i is μ_i and their variance is σ_i^2; then the simple "standardization" **z** such that $z_i=(x_i-\mu_i)/(\sigma_i\sqrt{N})$ can easily be decoded from this population, and **z** will have components of expected value 0 and variance 1/N – see computation in note below[7]. This argument lends support to the idea that the implausibility of standardization is a red herring for the question of the possibility of modal HRR representations. But the weakness of this argument is that, due to NEF's way of computing weights (by combining transformational decoding and re-encoding), standardization may be processed by every set of weights projecting to the population from another, and de-standardization from every set of weights projecting from the population to another. We will come back to the question of standardization, but for the moment we will proceed as if our theoretical argument was correct and the implausibility of standardization was indeed a red herring.

Generation of structured representations and decoding.

The structured representations that we use in these experiments are simple lists in which the *what* representation for a randomly chosen image is convolved with the *where* representation for that same image[8]; a given number (1 to 8) of such convolutions is generated and then summed:

$$\mathbf{s}=\Sigma_{i=1}^{N}\,\mathbf{what}_i*\mathbf{where}_i$$

The sum is normalized, and then convolved with the approximate inverse of one of the element of the first term of the sum - i.e., $\mathbf{p}=(\mathbf{what}_1)'$ or $\mathbf{p}=(\mathbf{where}_1)'$, giving $\mathbf{r}=\mathbf{s}*\mathbf{p}$. The best match is found by calculating the dot product of the result **r** with each of the 2x81 representations in the lexicon; we then compare the result to the other member of the first pair to see if the decoding is correct. In the "mixed" case, each pair is constituted by binding a modal *what* representation to a random representation, and the task is to find the random representation given the modal one, as is the case in Hunsberger et al.'s (2013) experiments, although here all 2x81 basic representations are searched for a match, not just random ones.

counterintuitive, constant terms can be decoded (approximately and over a finite input range) from a population of a non-constantly activated neurons, as a few minutes spent playing with nengo (nengo.ca) will show.

7 If the equation for the current of a neuron in the x-encoding population is $J = \Sigma w_i x_i+\beta$, we can set $w'_i = (\sigma_i\sqrt{N})w_i$ and $\beta'= \beta+ \Sigma w_i\mu_i$ so that the current of the neuron with the new encoding weights and bias is $J' = \Sigma\, w'_i z_i+\beta' = \Sigma(\sigma_i\sqrt{N})w_i(x_i-\mu_i)/(\sigma_i\sqrt{N})+ \beta + \Sigma\, w_i\mu_i = \Sigma w_i(x_i-\mu_i)+\beta+\Sigma\, w_i\mu_i = \Sigma w_i x_i+\beta = J$.

8 Obviously this way of naming things is really appropriate only to the learning case: the non-learning networks do not produce representations that are inherently of type *where* or *what* (just a neurally computed function of the image) and the random representations do not inherently represent anything.

Tests and controlled variables

For each type of representations (random, learning, no learning/random-modal, mixed, standardized as a whole) and for dimensionalities 16, 32, 64, 128, 256 and 512, we generated 100 sets of 2x81 representations and tested each number of summands from 1 to 8 one hundred times. We didn't do any statistical tests (with one exception, see the Analysis section below), but the numerical data in the graphs is in the Supplementary Material (Data-De-Pasquale-Poirier-2014.xls) along with the standard deviation for each average.

Results

We show the graphs for all aforementioned dimensionalities (full results are in the Supplementary Material)[9]. As can be seen on the graphs, at low dimensionality, the accuracy of every type of representations deteriorates rapidly as the number of summands grows to 8. For 16 and 32 dimensions, the difference is small between types of representations.

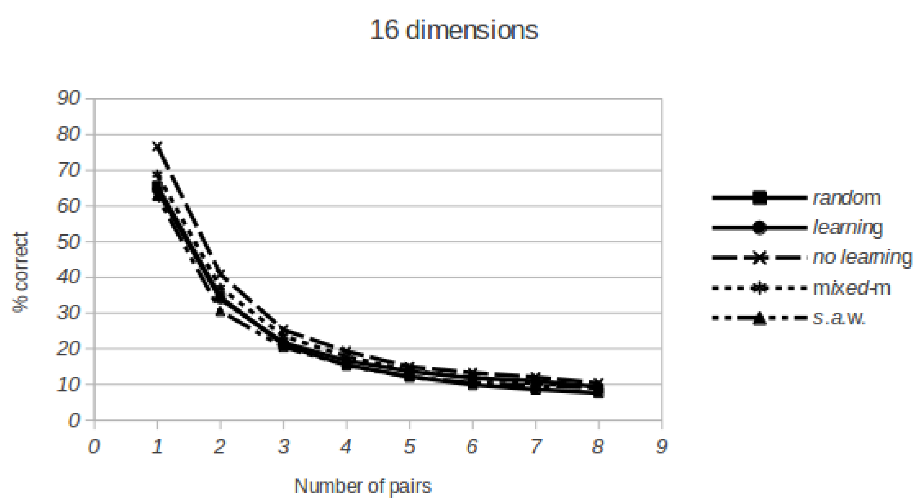

Figure 1 Test for 16 dimensions.

[9] Please note that we chose the range of the y axis so that differences between kinds of representations can be easily seen; the y axis does not always start at 0% and sometimes stop before 100%.

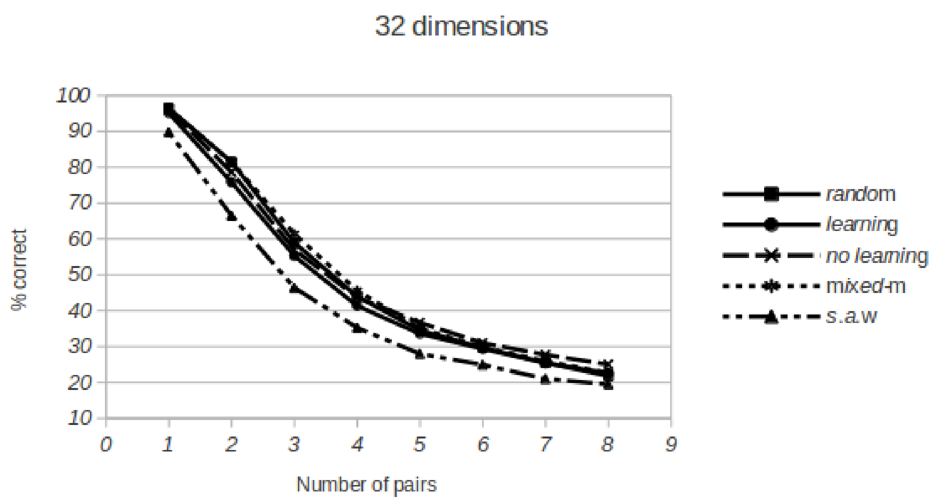

Figure 2 Test for 32 dimensions.

But as dimensionality increases, we observe that there is a distinct cost associated with using modal representations.

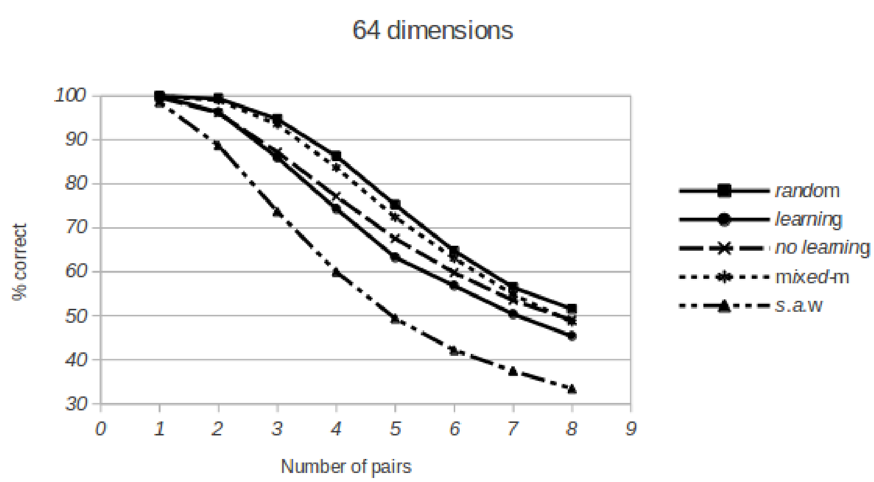

Figure 3 Test for 64 dimensions.

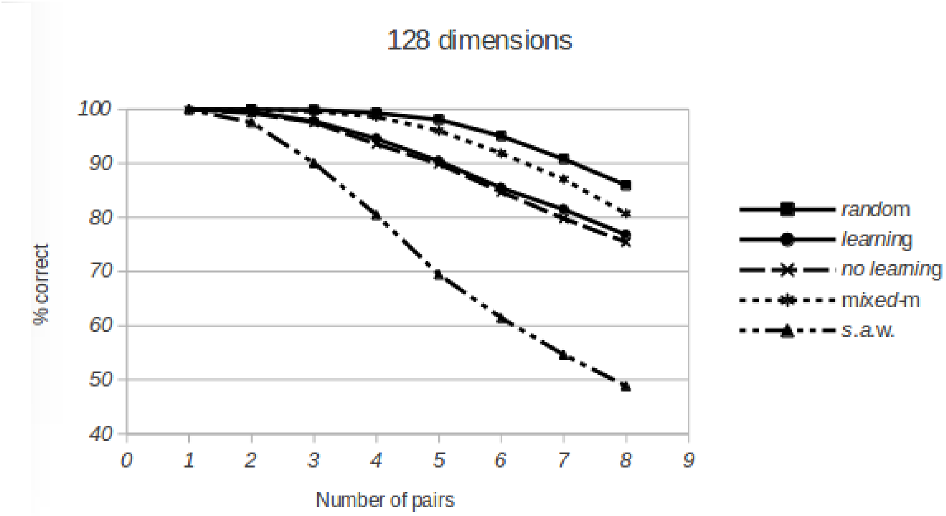

Figure 4 Test for 128 dimensions.

In fact, the types of representations cluster in three to four groups: Random representations perform at near perfect level. Learned modal and random-modal representations form another group that deteriorates in a noticeable but perhaps acceptable manner as the number of items in a sum grows, with "mixed" representations staying most of the time below random representations and above random-modal and learned ones, for experiments above 32 dimensions.

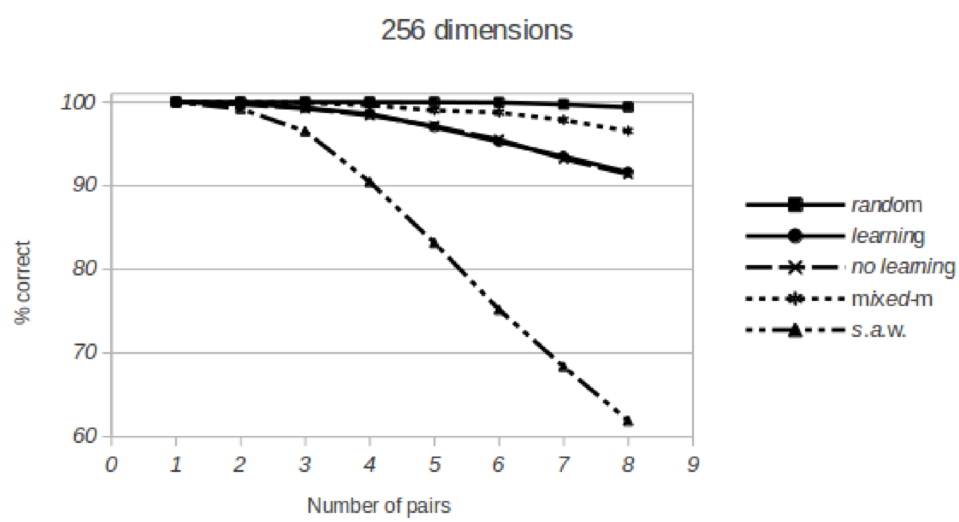

Figure 5 Test for 256 dimensions.

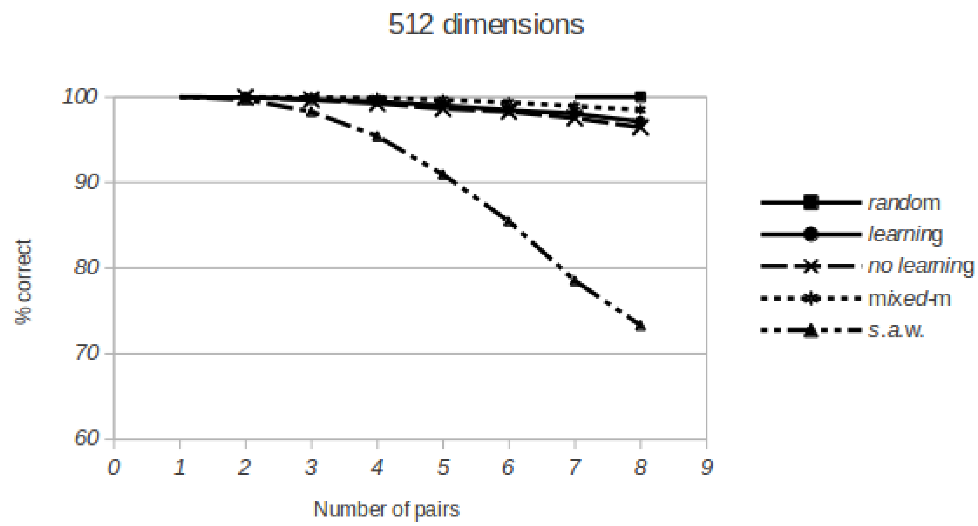

Figure 6 Test for 512 dimensions.

Finally, the set of modal representations that are standardized as a whole behave very poorly even in high dimensions, although deterioration is less pronounced in higher dimensions.

Notice that on the 8 pairs test, for 512 dimensions, we obtained 97.13% correct for modal representations and 96.45% correct for random-modal representations. Since the SPA generally operates, at the cognitive level, in high dimensionality (512 dimensions), it may be that the cost of modality is not so dire in such contexts.

Analysis

We were somewhat concerned that the "learned" representations were not really learned, given their results' similarity with those of random-modal representations. In high dimensionality, a correctly chosen basis of nonlinear functions can serve to approximate almost any function, and even a relatively randomly chosen basis can be useful (this is the idea behind RBF networks). So we trained a network on both tasks, freezing the input-to-hidden-units weights. Error did diminish more for high dimensional hidden layers, reaching very low levels for 512 dimensions, but still wasn't as low as with full learning, and this for both tasks[10]. We

[10] We ran Welch t-tests for samples of unequal variances, comparing 30 networks in the full learning condition and 30 in the frozen input-to-hidden-layer weights condition, for both tasks (number of dimensions: 512, number of epochs: 500, learning rate: 0.05, momentum: 0). The difference was highly significant ($p<10^{-10}$) in both cases ($t \approx -71.91$ in the

conclude that, especially for the *what* task, the hidden representations do contain task-specific information. The similarity in results on the sum of pair tasks may be due to the fact that both kinds of representations possess a correlational structure that is not present in random representations.

We also wanted to know more about standardization: since the way in which standardization is done changes drastically the performance of HRR operations (with representations standardized as a whole falsifying H2 only for the two highest dimensionalities considered), it would be interesting to know whether standardization is, in fact, necessary to apply HRR operations to modal representations. So we ran a test at 512 dimensions where no representations were standardized (except random ones, since the way they are generated make them automatically standardized). The answer to this question is that standardization is obviously necessary; we couldn't reject H2, with learned and random-modal representations producing near-zero accuracy. To be more precise, if we average for all numbers of pairs (1 to 8), learned representations are at 0.555% correct decoding and random-modal are at 0.67375% correct decoding; this close to what a random guess would achieve; as we said, a random guess would result in about 0.6% correct "decoding".

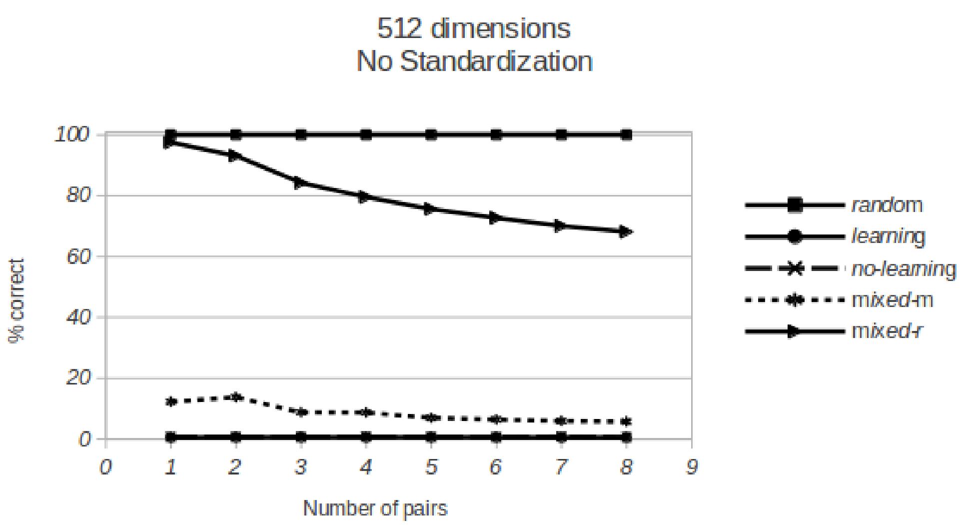

Figure 7 Test for 512 dimensions without standardization.

what case, with degrees of freedom approximately 29, t≈-66.37 in the where case, with degrees of freedom approximately 29.32). Since exploratory runs showed that the difference between full learning and learning only in the second layer of weights decreases sharply as the dimensionality of the hidden layer increases (as it would be expected to do), this seems to indicate that hidden layer learning does significantly impact performance at all the dimensionalities considered, and therefore that the learned representations must be somehow different from random-modal ones, even for high dimensionality.

By contrast, the decoding that involves mixed representations fared much better than a random guess. Mixed representations can even falsify H2. Interestingly, this is true only when the sum is convolved with the inverse of the random representation and the target representation is the modal one, which is the reverse of Hunsberger et al.'s experiment; in this case, the performance started at 97.48% for 1 pair and fell to 68.13% for 8 pairs. If the sum is convolved with the modal representation and the corresponding random representation (the label) is searched, as in Hunsberger et al.'s experiment, the results are a little better than with fully modal representations, but not by as much, and not enough to falsify H2 (starting at 12.24% for 1 pair and falling to 5.72% for 8 pairs). While the notions of shallow and deep semantics apply mainly to modal representations, random representations, such as the number representations in Spaun, are sometimes considered less "deep" than modal pointers generated through learning (Eliasmith, personal communication). If this is so, without standardization, it appears that mixed representations support going from shallow to deep semantics (from the label to the modal pointer (a process called "dereferencing"; Eliasmith 2013) but not the reverse.

Discussion

We now want to point out limitations, but also interesting consequences of our simulations.

First, an interesting consequence of our results is the fact that the effect of modality on HRR processing, at least for representations generated by one layer of neural treatment (and possibly feedback from a second output layer), can be explored, up to a point, by using the initial weights of the networks. This is interesting because it means that much (although obviously not everything) of what there is to know on the accuracy of HRR processing when using modal representations can be discovered without the lengthy process of learning appropriate weights. Because of this, it is possible to make large-scale tests which would have otherwise been very costly in resources (i.e., time). If one wants to escape an analysis done with random initial weights, one has to go *up* – i.e., in the higher layers of a deeper network. Deep nets are a growing field, and Eliasmith's (2013) modal pointers are in fact generated through such deep networks (Tang and Eliasmith 2010), as are Hunsberger et al.'s (2013) pointers in the categorization model.

Second, notice that, in our experiments, only in the case of the 16 dimensional representations, where all types of representations perform poorly, are the modal representations really "reduced" or "compressed" version of their input. It might be the case, and that would be something to be concerned about, that the small size of the input combined with the small size of the lexicon (2x81), relative to the dimensionality, is what accounts for most of the improvements we see on modal representations and that, accordingly, the performance of pointers that were the result of real compression would not improve as much. Further tests are needed to assess this possibility, perhaps with the MNIST dataset to allow for a higher-dimensional input. However, we doubt that these issues are fatal – it should be noted that it is already known that HRRs perform better in high dimensions (Plate 2003, Appendix D), so it is normal to expect good performance in high dimensions; notice also that the dramatic improvement is also present in random representations and cannot be explained by the relative size of the input and hidden layers, since those do not exist for random representation in this experiment.

A third important observation to make is that, without the 50% threshold, our experiment is not a decisive test of the more specific hypothesis (H2), only of (H1). While there is without a doubt a deterioration of decoding for modal representation, this deterioration is relatively minor for high dimensional representations. Whether the deterioration is judged to fall within acceptable boundaries, or to call into question only the more extreme version of the semantic pointer hypothesis, is very much subjective at the moment. Only the inclusion of such modal representations in a model that contained a large portion of the Semantic Pointer Architecture could really count as a strong test of H2; and even a failure of such a model would only be a failure of the semantic pointer concept interpreted 'imperialistically', that is, as the basis of all, or most, of cognitive processing. If non-modal concepts were to do most of the cognitive processing, the SPA could be fine even if such an "imperialistic" model were not successful[11]. But clearly, if we conform to our operationalization of H2, H2 should be rejected: at high dimensionality (128 to 512), properly standardized modal representations are decoded with accuracy well above 50% for the 8 pairs decoding task.

Fourth, it is clear that proper standardization is crucial: not only are HRR operations simply impossible with high dimensional, non-standardized modal representations, as our test in the Analysis section shows, but modal representations of a given type must be standardized separately from modal representations of a different type. When both sets of modal representations are standardized as a whole, for 8 pairs, decoding falls below 50% except at the highest two dimensionalities considered (256 and 512), and even then it is very poor. The

[11] It is not clear whether Eliasmith would be "imperialist" by this definition; although he is open to the possibility of amodal pointers (see chapter 10 of Eliasmith 2013), he is also committed to solving the grounding problem and his solution involves modal pointers.

standardization process must respect the structure of the set of representations: the heterogeneity of the set must not be too great. It is not sufficient that, as a whole, the modal HRRs have a mean of zero and variance 1/N: it is necessary that *each* subset belonging to a *distinct* type has a mean of zero and variance 1/N. A set of representations must not be too heterogeneous if we need to apply the standardization procedure. If the need for standardization is a red herring, as our theoretical argument involving NEF's characterization of representations would suggest, then it is not important to consider such matters. But if our theoretical argument is ultimately not convincing, and we do want to hold on to the idea that there are modal HRRs in the brain, then we must consider seriously the hypothesis that the process of standardization is real and happening for each different kind of ways there is of generating HRRs; for instance, HRRs for digits will be standardized separately from HRRs for other kinds of images, and these would be standardized separately form HRRs for sounds.

Finally, it must be noted that, though largely qualitative, our test is more strict than Hunsberger et al.'s (2013) as a test of the viability of modal HRRs (their goal was not, admittedly, to test the viability of modal HRRs, but their experiment does provide some relevant results for such a test). Indeed, "mixed" sets of representations, where the pairs each comprise one modal and one random patterns, generally (i.e., for all tested dimensionalities except 16 and 32 dimensions) perform less well than random representations but better than modal ones. Moreover, the mixed representations are the only ones that work well enough to falsify the H2 hypothesis when no standardization occurs (although in the task that is the reverse of the one considered on other tests). This validates the motivations behind the present experiment: only by testing a system of representations that is purely, or at least mostly modal, can we learn about the viability of the idea of modal HRRs, and, at least in its grounded incarnation of the semantic pointer idea.

Conclusion

We began by recalling that early assessments of convolution-based representational schemes saw them as unfit for use with perceptually-generated representations, and that even Plate called for mappings between such modal representations, on the one hand, and the more mathematically constrained HRRs, on the other, to interface the perceptual and cognitive systems. We then introduced Eliasmith's novel idea of semantic pointer, which precisely combines HRR processing with modal representation; Eliasmith claims that, if done the right way, such a combination solves the grounding problem. But we then pointed out that there are few tests of the claim that this kind of combination is possible; Thagard and Stewart (2011) only test one-convolution decoding, and Eliasmith et al. (2012) map their modal representations to randomly generated HRRs so that they can be used in more cognitive processing. Only one paper we could find, Hunsberger et al. (2013), truly tests this claim, albeit indirectly and partially, using a mix of random and modal representations. Our experiment directly tests the hypothesis and is more general: it extends the test to fully modal and random-modal representations and also give results for mixed representations and random ones, allowing us to compare the performance of all these different kinds of representations on the task considered.

We therefore proceeded to test the claim that modal representations cannot be processed well by HRR operations – and that the level at which this was true prevented the semantic pointer idea to work at all. We did so by comparing randomly generated representations with two kinds of modal representation (generated from images) and, to compare our work to Hunsberger et al.'s, with mixed representations. The results show that there is a distinct cost in accuracy of HRR processing for modal representation, but that this cost diminishes in higher dimensions. The cost is very high if standardization is done on a heterogeneous set of representations, but much lower if the structure of the set is respected. "Mixed" pairings, where a modal representation is paired with a random one, are easier to decode than pure modal pairings, and outperform fully modal representations when no standardization occurs (slightly or strongly depending on what computational role is played by which kind of representations). It remains to be seen if the fact that the costs of using modal representations diminish in higher dimensions is an artifact of the low cardinality of the lexicon used in our task, and if the diminished accuracy is tolerable for cognitive modelling purposes. Because of that, we cannot pronounce ourselves definitively on the viability of fully semantic pointers, though the results certainly allow us to be hopeful. But we do remind the cognitive modelling community that modal HRRs are a different kind of beast compared to randomly generated ones, and that the associated problems will have to be solved by anyone claiming to uphold the semantic pointer hypothesis.

References

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, 338(6111), 1202-1205.

Fisher, A. D., Lippincott, W. L., & Lee, J. N. (1987). Optical implementations of associative networks with versatile adaptive learning capabilities. *Applied Optics*, *26*(23), 5039-5054.

Hinton, G. (1990) Mapping Part-Whole Hierarchies into Connectionist Networks. Artificial Intelligence 46, 47-75.

Hunsberger, E., Blouw, P., Bergstra, J., & Eliasmith, C. (2013) A Neural Model of Human Image Categorization. In *35th Annual Conference of the Cognitive Science Society*} (pp. 633-638).

Markman, A., & Stilwell, H. (2004). Concepts á la modal: An extended review of Prinz's Furnishing the mind. Philosophical Psychology, 17(3), 391-401.

Plate, T. 2003. Holographic Reduced Representations, Stanford, CA: CSLI.

Rueckl, J. G., Cave, K. R., & Kosslyn, S. M. (1989). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of cognitive neuroscience*, *1*(2), 171-186.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. Artificial Intelligence, 46:159-216.

Stewart, T. C., Tang, Y., & Eliasmith, C. (2011). A biologically realistic cleanup memory: Autoassociation in spiking neurons. Cognitive Systems Research, 12(2), 84-92.

Tang, Y., & Eliasmith, C. (2010). Deep networks for robust visual recognition. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 1055-1062).

Thagard, P., & Stewart, T. C. (2011). The AHA! experience: Creativity through emergent binding in neural networks. *Cognitive science*, *35*(1), 1-33.