Paired Associates Testing: Addressing Retroactive Interference with the Semantic Pointer Architecture

A Report Submitted in Partial Fulfillment of the Requirements for
SYDE 750

Matthew Robichaud
20532131

Faculty of Engineering
Department of Systems Design Engineering

April 20, 2020

Course Instructor:
Andreas Stöckel

UNIVERSITY OF
WATERLOO

Contents

1 Introduction

A classic experiment in psychology, first described in 1894 by Calkins is the A-B, A-C paired associates test. In the test, the association between a set of pairs is learned, 'A' paired with 'B', followed by a second set of pairs. This experiment explores the psychology of association and interference between newly formed memories.

In paired associates experiments, familiar objects are used to form the associations, ranging from associating a given colour to a number, two nouns together, or a subject and a verb. The most prominent effect of the A-B, A-C test is that learning of the second set impairs recall of the first set, such that when presented with the cue (A) more of the items from list C are recalled than list B.

Two types of memory interference have been identified: proactive and retroactive interference. In the former, older memories can inhibit new ones from forming, but more commonly, new memories (associations) lead to the forgetting of older ones, called retroactive interference.

An important debate surrounding the idea of memory interference is the actual cause of forgetting: with one possible explanation being competition between the memories, and another being associative *unlearning*. Due to competitive interference of similar memories, recency effects will favour the newer pairs (Postman and Stark, 1969). Associative unlearning on the other hand postulates that the new associations cause unlearning of the initial pairs, effectively replacing them in memory (Barnes and Underwood, 1959). Although overall competition seems to be favoured, recent research suggests these ideas describe what takes place only before memory consolidation, and cannot explain findings when older memories are probed (Wixted, 2004).

Between these two explanations for the cause of forgetting in human subjects, the unlearning effect is readily observed in artificial neural networks. Here, new inputs are known to catastrophically interfere with previously learned connection weights. (McCloskey and Cohen, 1989). Generally, learning to process new inputs superposes information on the previous learned connections, thus destroying them.

Semantic pointers provide a framework for performing computation using symbolic representations. Unique to the semantic pointer architecture (SPA), described by (Chris Eliasmith, 2013), is the grounding of those symbols in experience through the compression of higher dimensional neural representations.

Auto-association is a key component of the SPA due to its role in clean-up after algebra has been performed with pointers. Furthermore, a method was developed to demonstrate the capability of Nengo models in performing heteroassociative tasks. This implementation uses a learning rule to drive the decoded representation of the input keys towards their paired values (Voelker, E. Crawford, and C. Eliasmith, 2014).

The SPA enables multiple representations to be compressed into one and later retrieved. In principle, this should allow multiple associations to be learned nondestructively. The objective of this work is to explore the ability of semantic pointers to limit retroactive interference in a paired associates learning task.

2 Neurobiological Model

2.1 System Description

Certain characteristics of the A-B, A-C paradigm should be included in the model. First, different stimuli (*i.e.* A_1 and A_2) should be learned and recalled independently, as it is assumed they do not have an affect on one another's associations. The number of pairs in a typical paired associates test with human subjects ranges from 7 to 12 pairs per set. When there are too many pairs, recall is impaired and this assumption breaks down. Generally, even when there are few pairs, the participant accuracy in the control condition is not 100%.

The conditions of the paired associates test are as follows. In all cases, the pairs are learned to a certain level of accuracy during each rehearsal, or learning, phase.

i) The rest group learns the A-B pairs, then rests for a period of time

ii) The control group learns the A-B pairs, then learns pairs C-D consisting only of new items.

iii) A third group learns the A-B pairs, then learns a list of pairs A-D.

iv) Other groups may in the second learning phase learn A paired with a reordered list of B, B paired with C, or other variations.

The groups are then presented with items A as cues and asked to recall the associated items. There are several ways the recall task can be implemented. One consists of a free-recall, where the cues are all presented at once and participants can list as many pairs as possible. Another is a cued-recall, where cues are presented one-by-one.

Generally, time is not entered into the experiment as a variable, and the recall task is performed shortly after the second phase is learned. The most prominent finding is the degree of retroactive interference between C and B in the third group, and the other conditions serve as controls of various other memory phenomena.

2.2 Design Specification

To implement a neurobiological system that stores persistent information about associations, its architecture must support two tasks: learning to associate a stimulus with a paired input, and being able to recall the paired item when prompted only by the stimulus. Only the stimuli should produce a valid response when used as cues.

10 items are used in each pair to be representative of the list sizes used in human experiments.

Each pair is learned for a period of 0.2 seconds before moving to the next pair. Once a set of pairs is learned, it is rehearsed one time without learning. The network then proceeds to learn the next set of pairs.

The items used for paired associates tasks can be thought of as high dimensional symbolic representations. They must be sufficiently dissimilar from one another in order to discriminate between them. For this implementation, 32 dimensions are used for the semantic pointer vocabulary. 200 neurons are used per dimension.

2.3 Implementation

To implement the paired associates test, a network was designed that uses the same basic architecture during the learning and recall tasks. The inputs to the network are the stimuli, the desired responses, and the learning signal. The learning signal controls the phases of the test (learning or recall) by inhibiting changes to connection weights through the error population.

The stimuli are processed by the vector Oja learning rule described by (Voelker, E. Crawford, and C. Eliasmith, 2014). The decoders of that layer are updated according to the prescribed error sensitivity (PES) method to output the bound representation of the stimulus and response for each stimulus. Because the encoders are set by the vector Oja learning rule, the stimuli are capable of learning independently.

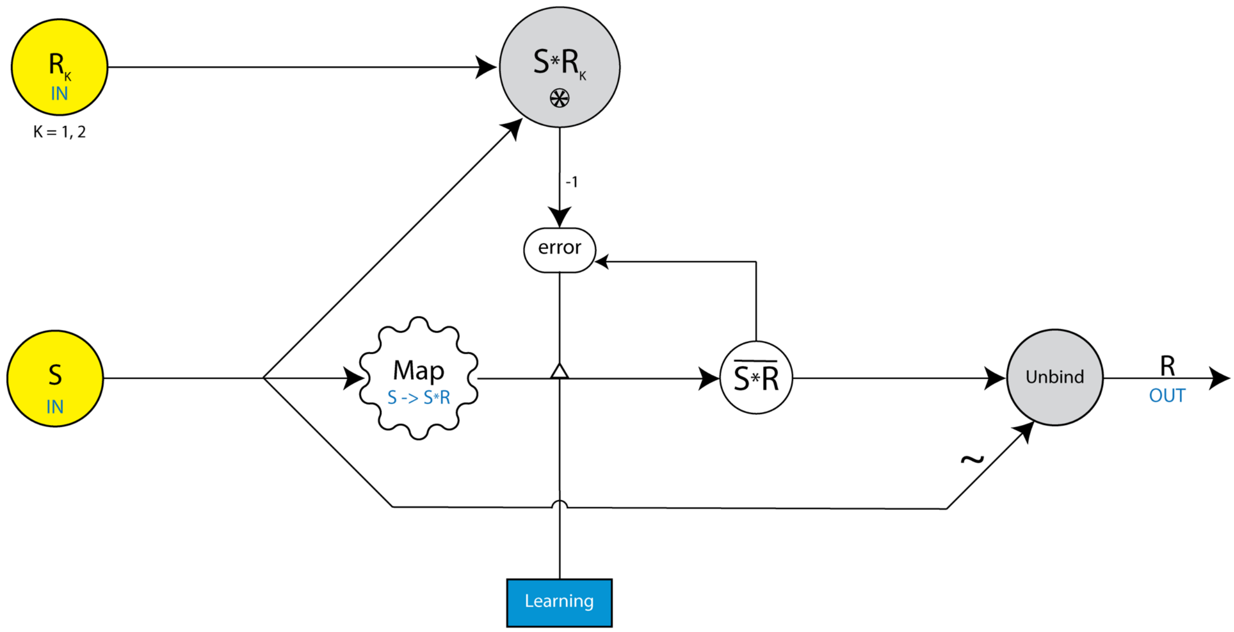

Figure 1: The architecture used for the paired associates learning task. During the learning phase, S contains stimuli and during the recall phase S is used as the cue for recall. R_{IN} is used during learning only. Test phases are controlled by the learning signal, shown in blue, which takes on the values 0 or -1. The other network nodes are neural ensembles. Grey circles represent pre-built circular convolution networks. The Map ensemble learns the vector Oja encoding and decodes the bound representation according to PES and the difference stored in the $error$ ensemble. The bound representation is stored in $\overline{S \circledast R}$, unbound using the inverted cue, and output by the network.

The system is implemented largely in Nengo (Bekolay et al., 2014), not Nengo_SPA, but uses vectors generated from the SPA library. Results are shown in terms of similarity between outputs and objects in the SPA vocabulary. Each experiment uses a new randomly generated vocabulary.

The system illustrated in Figure 1 is subjected to a paired associates learning and recall test with learning controlled by the inputs and the learning signal, which inhibits learning except for in the first phase, where the first set of pairs are learned, and the third phase, where the second set of pairs are learned, shown in Figure 2.

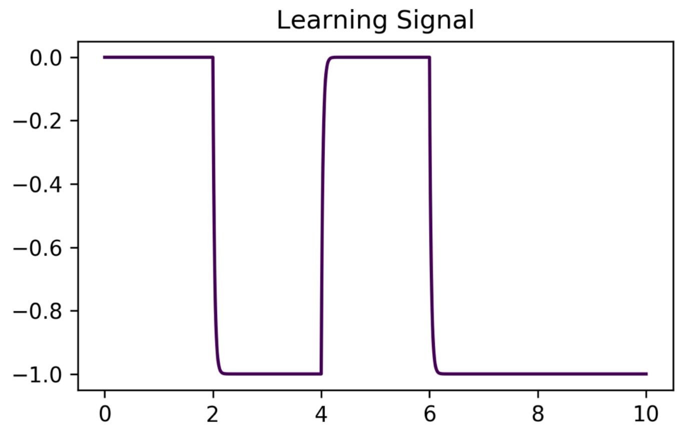

Figure 2: The learning signal inhibits learning when negative and has no action otherwise.

In order to test retroactive interference in the final phase, the first list of stimuli are shuffled to generate cues and appended to the list of stimuli.

3 Experiments and Results

3.1 A-B, C-D Control Test

The control group of the A-B, A-D experiment should demonstrate minimal potential for retroactive interference. By training the second list of pairs with different stimuli than the first list, the vector Oja rule employed in the learning model should prevent the initial pair associations from being overwritten.

The stimuli and desired responses for A-B, C-D learning are shown in Figure 3.

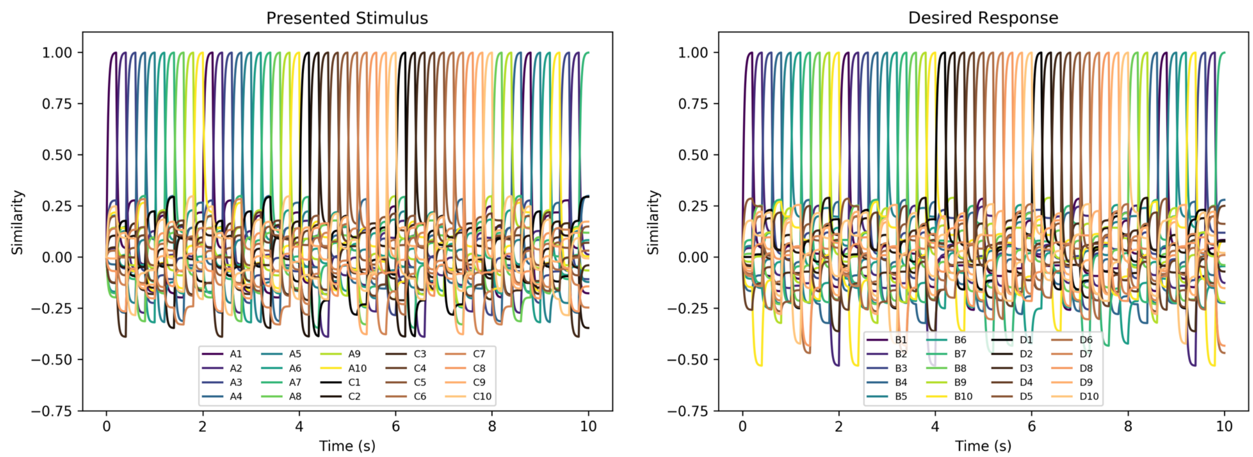

Figure 3: The stimuli and responses for the A-B, C-D experiment shown. The elements from the lists A and B are shown in one set of colours and the elements from lists C and D in a second set (orange).

Once processed by the input layer of neurons the stimulus is mapped to the associated bound pair. The mapping is successful and the decoded representation shows high similarity to the ground truth and discriminates well between the different pairs (Figure 4). This again indicates the vector Oja rule has been successful at recruiting a subset of the layer's neurons for each stimulus.

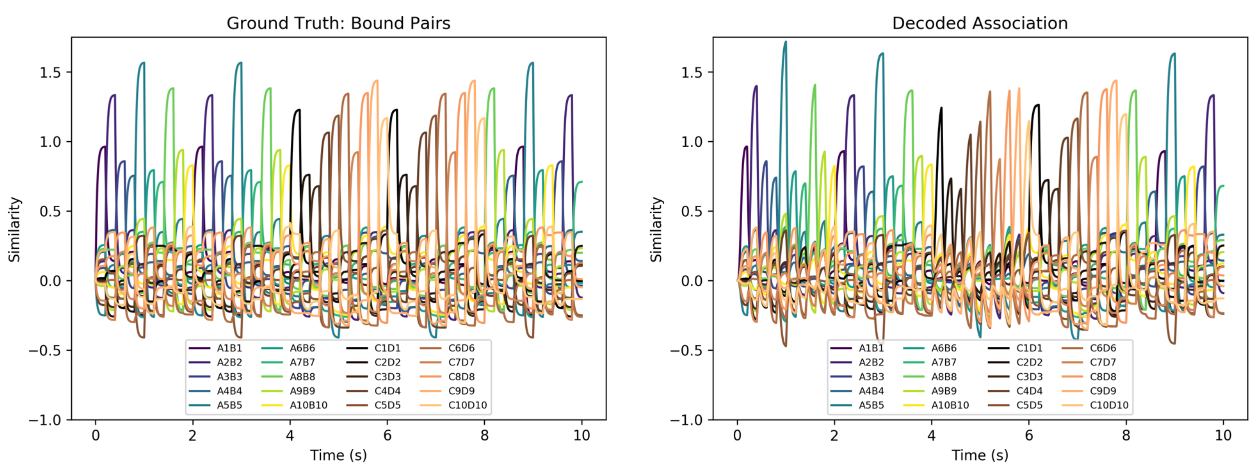

Figure 4: For the A-B, C-D control test shown left is the ground truth symbol binding and right is the actual learned association.

The output is then processed by unbinding the pair with the stimulus. As can be seen in Figure 5, the paired items from the first learning task are recalled accurately. The order is different from learning as the cues were shuffled. Accuracy in this control test was 100% as expected.

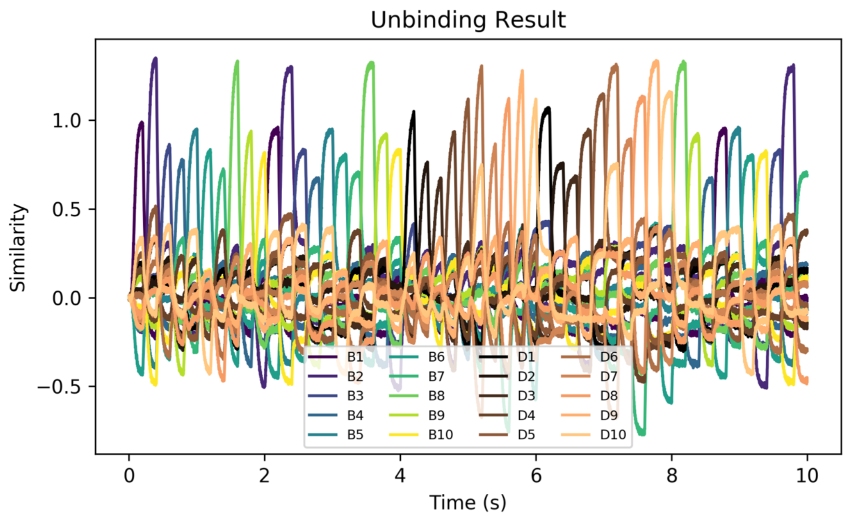

Figure 5: The result of the A-B, C-D control test shows accurate recall of the first pairs after learning an unassociated list of pairs.

3.2 A-B, A-D Test

The test condition which demonstrates retroactive interference in human participants is the A-B, A-D paired associates. The inputs are shown in Figure 6 below. The desired responses are identical to the first experiment but the stimuli are all from the A series.

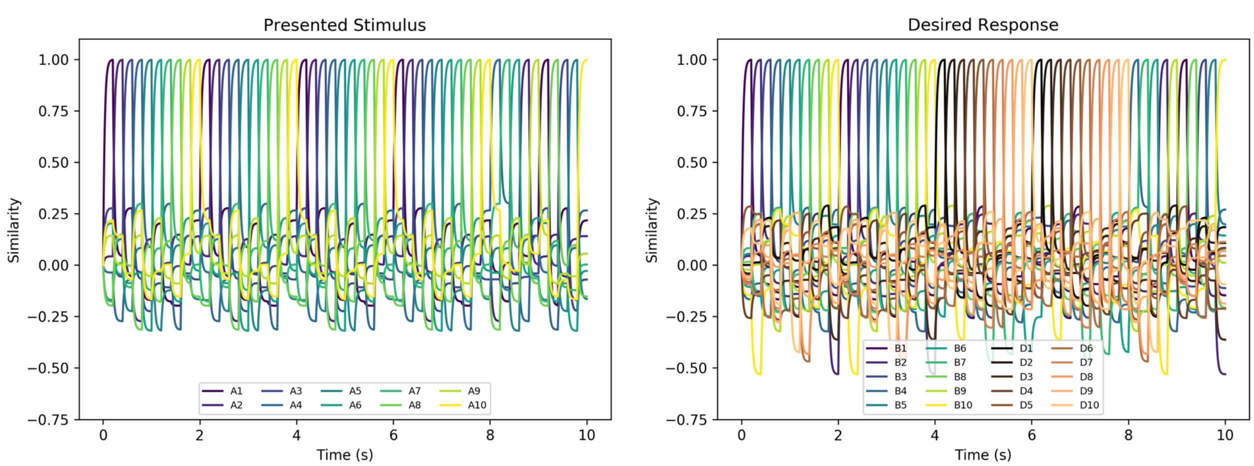

Figure 6: The stimuli and responses for the A-B, A-D experiment shown. The elements from the lists A and B are shown in one set of colours and the interfering elements are shown in a second set (orange).

In this case, because the same stimuli are used for both associations, associations to items B are not recovered in the test phase. This demonstrates retroactive interference (Figure 7). Here, the vector Oja rule alone fails to preserve the memory of previous associations.

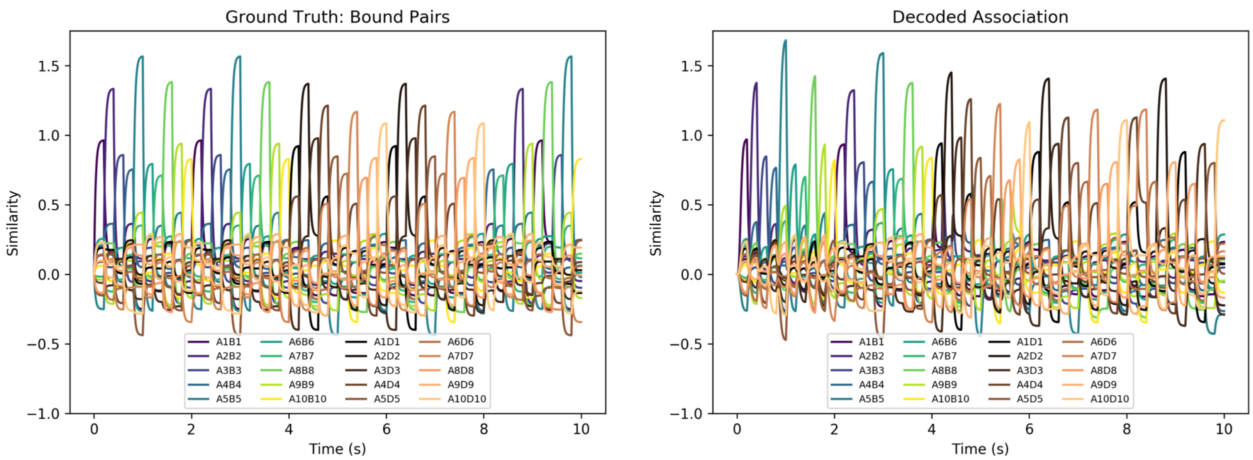

Figure 7: The A-B, A-D ground truth and actual learned association are shown.

As expected, once the decoded pairs are unbound, only items from list D are recovered (Figure 8). The recall during training was 100% (20/20) but in the final test, none of the items from list B were recalled.

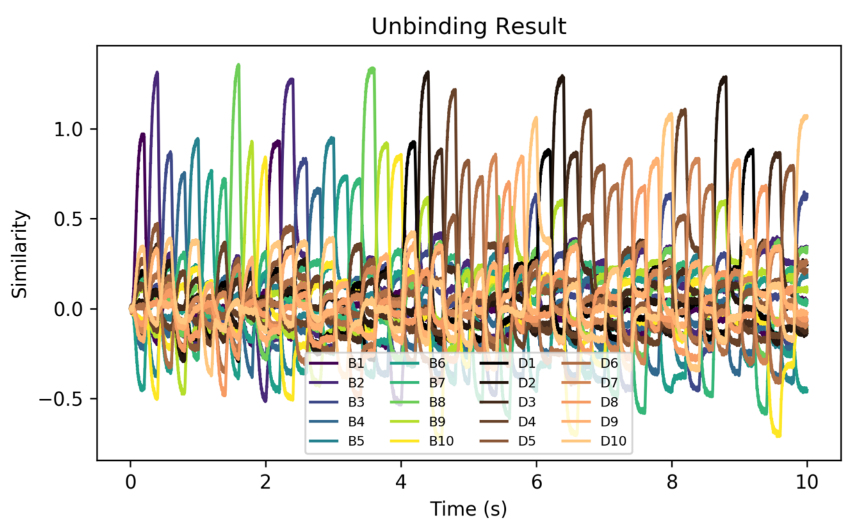

Figure 8: In the final test after A-B, A-D learning, items A are used as cues and only elements from list D are recalled.

3.3 A-B, A-D with Superposition

To examine the effect of pointer superposition in the learning phase on the recall of items B, a last experiment was performed. Initially, the network was modified to include a memory ensemble which stores previously bound representations and feeds them to the error node. This was unsuccessful however because it interfered with the Oja encoding and only acted to deteriorate the quality of learning. This will be discussed in more detail in the next section.

Instead, to simulate "memory" of the previously learned pairs, the ground truth was modified such that the superposition of the two associations would be learned:

$$\mathrm{GroundTruth} = A \circledast B + A \circledast D$$

Figure 9 shows that both pairs were stored and recalled in contrast with the previous experiment.

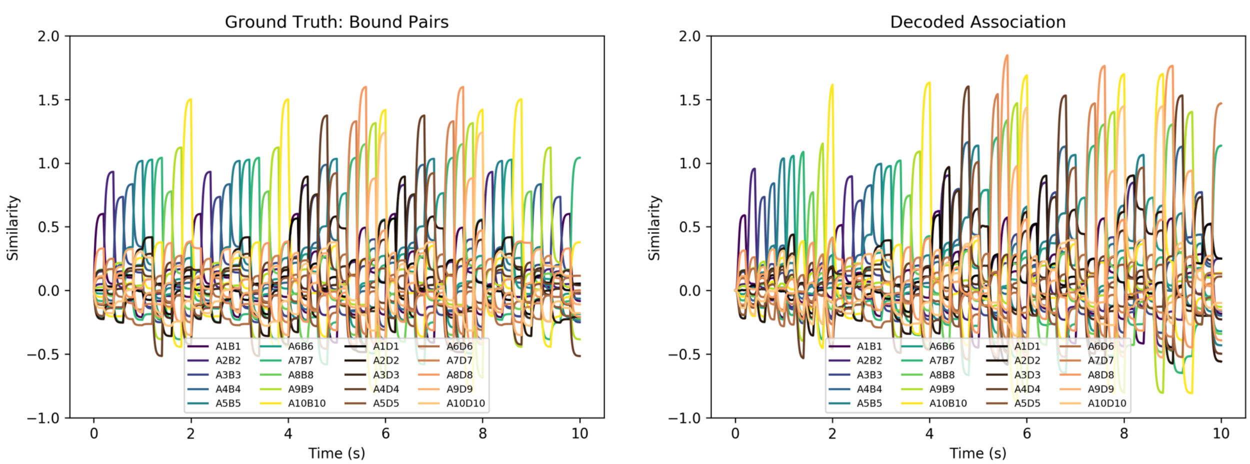

Figure 9: The A-B, and superposed A-B + A-D ground truth and actual learned association are shown.

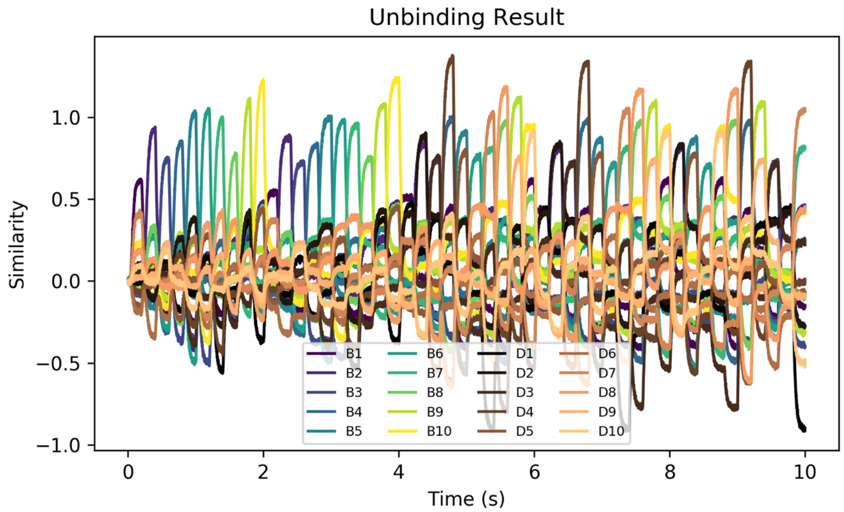

Figure 10: In the final test after A-B, A-D learning, items A are used as cues and a superposition of elements B and D are recalled.

When unbound through the circular convolution layer, both items from lists B and D were recalled (Figure 10). The accuracy of the recall is summarized in Table 1. It can be seen that learning of the second list was slightly impaired by superposing the first bindings. The recall of items A was strong however, with 98% accuracy over 5 experiments (only once was the corresponding B item not in the top 2 most similar vectors).

Table 1: The recalled vectors were analyzed for their similarity to vocabulary elements in response to the stimuli (10 items). For each time the output was most similar to the correct item, "most" was incremented by one. Percent correct refers to the mean accuracy when the top two most similar elements are considered. Data are means of N=5.

Similarity	First List	Second List	Test
Most	10	4.4	4.4
Second-most	0	3.8	5.4
Top 2	10	8.2	9.8
Percent	100%	82%	98%

4 Discussion and Conclusion

The vector Oja rule is a convenient way of representing multiple independent objects in a single network. It accomplishes this by taking advantage of the diversity of neural tuning curves.

Supporting independent associations by the vector Oja encoder learning rule enabled independent learning of pairs with differing stimuli (Section 3.1). This reproduced expected behaviour in the control condition. When the network was trained with A-B followed by C-D, it remained able to recall items from list B when cued with items from A. This rule alone however broke down when pairs were provided with the same stimulus, as in the case of A-B, A-D learning (Section 3.2).

In 3.3, superposition was used to eliminate retroactive interference without increasing complexity of the system or memory requirements. In these simulations, the superposed binding was learned directly from the ground truth, but future work could expand on this by incorporating a separate feedback mechanism. Instead of superposing after convolution, the list items could be superposed by feeding the learned response back into an additive node before convolving with the stimulus. This takes advantage of the distributive property of circular convolution (Eric Crawford, Gingerich, and Chris Eliasmith, 2016).

$$A \circledast (B + D) = A \circledast B + A \circledast D$$

This may be a manufactured problem however, because in larger networks, especially those which include symbolically grounded representations of the familiar objects used in pair association tasks, these objects would be represented by separate networks in the brain. In this case, dynamical subsystems that have memory may not interfere with other paired associates such as was encountered in these simulations.

References

Postman, Leo and Karen Stark (1969). "Role of response availability in transfer and interference". In: *Journal of Experimental Psychology* 79.1 PART 1, pp. 168–177. ISSN: 00221015. DOI: 10.1037/h0026932.

Barnes, Jean M. and Benton J. Underwood (1959). ""Fate" of first-list associations in transfer theory". In: *Journal of Experimental Psychology* 58.2, pp. 97–105. ISSN: 00221015. DOI: 10.1037/h0047507.

Wixted, John T. (2004). "The Psychology and Neuroscience of Forgetting". In: *Annual Review of Psychology* 55.1, pp. 235–269. ISSN: 0066-4308. DOI: 10.1146/annurev.psych.55.090902.141555.

McCloskey, Michael and Neal J. Cohen (1989). "Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem". In: *Psychology of Learning and Motivation - Advances in Research and Theory* 24.C, pp. 109–165. ISSN: 00797421. DOI: 10.1016/S0079-7421(08)60536-8.

Eliasmith, Chris (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. New York: Oxford University Press. ISBN: 978-0-19-979454-6. DOI: 10.1093/acprof:oso/9780199794546.001.0001.

Voelker, A., E. Crawford, and C. Eliasmith (2014). "Learning large-scale heteroassociative memories in spiking neurons". In: *Unconventional Computation and Natural Computation*. DOI: 10.13140/RG.2.1.1382.1688.

Bekolay, Trevor et al. (2014). "Nengo: A Python tool for building large-scale functional brain models". In: *Frontiers in Neuroinformatics* 7.JAN, pp. 1–13. ISSN: 16625196. DOI: 10.3389/fninf.2013.00048.

Crawford, Eric, Matthew Gingerich, and Chris Eliasmith (2016). "Biologically Plausible, Human-Scale Knowledge Representation". In: *Cognitive Science* 40.4, pp. 782–821. ISSN: 15516709. DOI: 10.1111/cogs.12261.