Creating False Memories in Humans with an Artificial Neural Network: Implications for Theories of Memory Consolidation

Serban C. Musca (Serban.Musca@upmf-grenoble.fr)
Stéphane Rousset (Stephane.Rousset@upmf-grenoble.fr)
Bernard Ans (Bernard.Ans@upmf-grenoble.fr)

Psychology and NeuroCognition Lab - CNRS UMR 5105, University Pierre Mendes France (Grenoble 2),
1251 Avenue Centrale, BP 47, 38040 Grenoble cedex 9, France

Abstract

Building on the human memory model that consider LTM to be similar to a distributed network (McClelland, McNaughton & O'Reilly, 1995), and informed by the recent solutions to catastrophic forgetting that suppose memories are dynamically maintained in a dual architecture through a memory self-refreshing mechanism (Ans & Rousset, 1997, 2000; Ans *et al.*, 2002, 2004; French, 1997), we checked whether false memories of never seen (target) items can be created in humans by exposure to "pseudo-patterns" generated from random input in an artificial neural network (previously trained on the target items). In a behavioral experiment using an opposition method it is shown that the answer is yes: Though the pseudo-patterns presented to the participants were selected so as to resemble (both at the exemplar and the prototype level) more the control items than the target items, the participants exhibited more familiarity for the target items previously learned by the artificial neural network. This behavioral result analogous to the one found in simulations indicates that humans, like distributed neural networks, are able to make use of the information the memory self-refreshing mechanism is based upon. The implications of these findings are discussed in the framework of memory consolidation.

Keywords: distributed information; neural networks; human memory; representation format in human memory; memory self-refreshing; exemplar theory; prototype theory; indirect memory test; familiarity; perceptual fluency.

Introduction

Can information be transported between a GDN–a multi-layered network trained by a gradient descent learning procedure–and humans? This question is central to models of human memory that suppose that LTM is similar to a distributed network (e.g. McClelland, McNaughton & O'Reilly, 1995) and that memories are dynamically maintained in a dual architecture through a memory self-refreshing mechanism (Ans & Rousset, 1997, 2000; Ans *et al.*, 2002, 2004; French, 1997).

GDN's memory gradually emerges as a result of the processing of the training exemplars: the connection weights between the processing units reach values that allow the network to perform correctly. Thus the memory of a given trained GDN can be conceived as the particular set of connections weights between its processing units.

Of course, when trained on a new set of exemplars S_2, the connection weights of a network previously trained on a set of exemplars S_1 change in order to allow the network to perform correctly on S_2, and this new connection weight set does not allow the network to perform correctly on S_1 any more (catastrophic interference or catastrophic forgetting: McCloskey & Cohen, 1989; Ratcliff, 1990). To get round this obstacle, an obvious solution is to train S_1 and S_2 concurrently, thus transforming sequential learning (i.e. <u>first</u> S_1, <u>then</u> S_2) into concurrent learning (i.e. S_1, <u>and</u> S_2 at the same time). However, concurrent learning relies on the assumption that S_1 is still available when S_2 is to be learned, an unreasonable assumption when GDNs are used to simulate human memory phenomena: Every old exemplar is not available nor is it learned anew each time some new exemplars are learned (Blackmon *et al.*, 2004). A first step towards a more plausible solution to the problem of catastrophic forgetting in GDNs in sequential learning tasks is due to Robins (1995): Once a network has been trained on S_1, its memory is sampled, thus generating "random input-computed output" pairs (or pseudo-items) that are stored in a non-neuromimetic memory; then, instead of training the network on S_2 only, it is trained <u>both</u> on S_2 and on the stored pseudo-items. If this solution makes it possible to reduce catastrophic forgetting in the absence of S_1 exemplars, it also resorts to an implausible "copy-paste" procedure in order to store the pseudo-items before they are used as training material.

The next solutions (Ans & Rousset, 1997; French, 1997) avoid the "copy-paste" procedure by having recourse to GDN architectures that are able to learn "on the fly" the "random input-computed output" pairs. For instance, Ans & Rousset's (1997) architecture is made of two separate GDNs, NET1 and NET2; once trained on S_1, NET1 generates reverberated "random input-computed output" pairs (called *pseudo-patterns*, PPs) that are used to train NET2. Then, when NET1 is to learn a new set of exemplars S_2, NET1 is not only trained on S_2 but also on PPs generated this time in NET2 (and conveying information on S_1). Were a third new training set S_3 to be learned, NET1's memory would first be transmitted to NET2 through PPs, then NET1 would be trained on both S_3 and PPs generated in NET2 (now conveying information on both S_1 and S_2). To sum up, this architecture is very efficient in avoiding catastrophic forgetting in the absence of old training exemplars and even in the case of numerous phases of sequential learning.

It will be shown that the memory of a GDN auto-associator can be transported to humans by means of PPs. However, this would not be a new result if the PPs presented to the participants amounted to the exemplars used to train the network, under their initial form or as noisy, distorted versions of them; except for the use of a neural network to generate the stimuli, this would be a classic learning/memory result (cf. Medin & Schaffer, 1978; Nosofsky, 1986; Posner & Keele, 1968; Reed, 1972).

The originality of this research is to show that a network's memory is transported to human participants and affects their behavior even though all the PPs used as stimuli are very different from the initial training exemplars.

In order to make this point, an opposition method was adopted: The PPs presented to the participants were selected so as to resemble more some other exemplars (hereafter *control items*) than the exemplars used to train the network (hereafter *target items*). Participants were only presented with selected PPs during a first, training phase, and then tested with the target items and the control items. A higher familiarity for the target items than for the control items–though the latter resemble more the PPs previously presented to the participants–would show a memory transport between the neural network and the participants.

The experiment is preceded by a simulation: A new (untrained) GDN auto-associator was trained on the very PPs stimuli used in the behavioral experiment. Starting from a comparison of behavioral and simulation results, the discussion will consider the possible interpretations of the behavioral data.

Simulation

The simulations involved the construction of 106 items, also used in the behavioral experiments (cf. Figure 1). The items were matrices constructed as follows. Starting from the centre of a 19×19 black grid, the following procedure was applied 20 times: A direction (up, down, left or right) was randomly chosen and two squares in that direction were turned white, then the last square served as starting point for the procedure on the next step. Any resulting pattern wider or higher than 13 squares was discarded, the remaining were re-centered on a 13×13 grid until 106 different and meaningless items were available. Each item was then coded as a vector of length 169, with black squares coded *0* and white squares *1*. This set of items was subsequently ***randomly*** divided into two lists of 53 items, List A and List B.

The opposition method outlined in the introduction was used in the simulation. A first network (NET1) was trained on the target items (e.g. List A items), then generated PPs that were selected so as to retain only those that resemble more the control items (i.e. List B items) than the target items. Then, a new network (NET2) was trained only on the selected PPs and tested on the target and control items. The opposition method allows for the following contrasted predictions.

If the selected PPs are but distorted items, owing to the selection constraints (exposed below), they are distortions closer to the control items than to the target items; thus NET2, trained only on the selected PPs, would exhibit a better performance at test on the control than on the target items. On the contrary, it may be that despite the constraints applied to the PPs, the selected set of PPs convey enough important information on the function instantiated by NET1 once it has been trained on the target items for NET2 to instantiate a similar function; in this case NET2, trained only on the selected PPs, would exhibit at test a better performance on the target than on the control items.

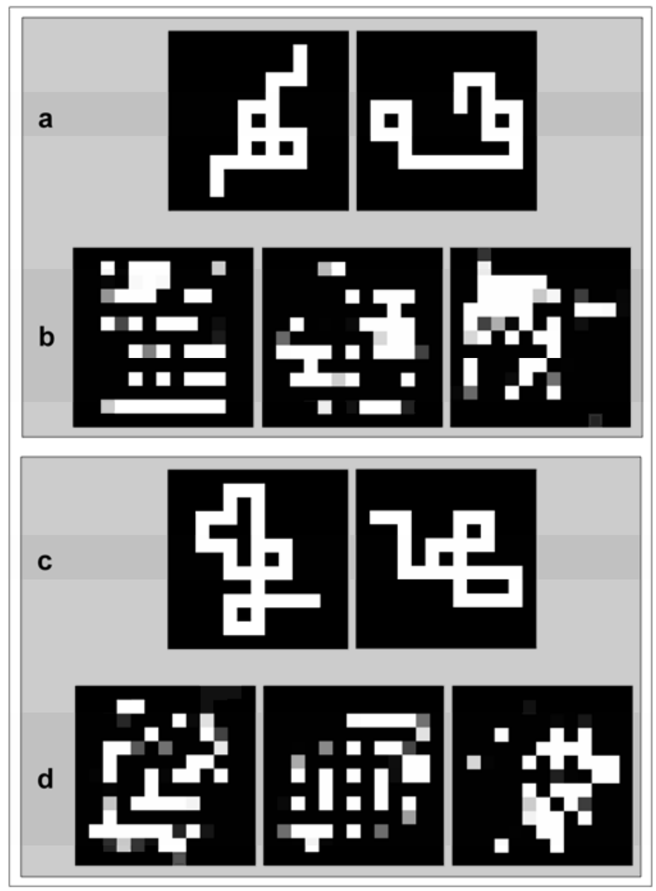

Figure 1: Examples of the experimental material: a) List A items; b) PPB_A pseudo-patterns; c) List B items; d) PPB_B pseudo-patterns. See text for details.

Material and Procedure

NET1 is a backpropagation three-layer auto-associator with 169 input units, 169 output units, 16 hidden units, a learning rate of .01 and a momentum of .7, initialized with random connection weights uniformly sampled between -0.5 and 0.5. After NET1 has been trained on the target items, then PPs were generated according to Ans & Rousset's (1997, 2000) procedure: Binary random input was fed to the input layer, resulting activation propagated through the network to the output layer, the output was then fed to the input layer (*reverberation*), and again propagated through the network, and so on. PPs are output patterns produced after five re-injections (cf. Figure 1). Out of 4,325,000 PPs generated in this way, only those PPs that complied with all the three following rules were retained:

(R_1) In terms of Euclidean distance, each selected PP is closer to a control item than to any target item;

(R_2) To reduce the number of PPs while increasing their variety, the RMS distance between any two selected PPs is greater than .15;

(R_3) the mean of the Euclidean distances between each target item and the centroid of the selected base of PPs (the "mean PP", noted PP_M) is greater than the mean of the distances between each control item and PP_M ; with a formula:

$$\frac{1}{N}\sum_{i=1}^{N} d\left(T_i, PP_M\right) > \frac{1}{N}\sum_{j=1}^{N} d\left(C_j, PP_M\right) \qquad (1),$$

with T_i = target item i;

C_j = control item j;

$d(X,Y)$ = the Euclidian distance between vectors X and Y;

$$PP_M = \frac{1}{S}\sum_{k=1}^{S} PP_k \; ;$$

S = the number of selected PPs;

N = the number of target (control) items.

The significance of these selection rules is that they make sure that the PPs resemble more the control items than the target items, both at the exemplar level and at the prototype level. This is an important provision in the event of results showing a better performance at test on the target items. More precisely, without these constraints such a result could arise trivially, that is merely because the PPs would resemble more the target items (i.e. would be distorted versions of them).

When the target list for generating the PPs was List A, this procedure led to PPB_A, a 3000-PP base. For counterbalancing sake the 3000-PP base PPB_B was generated by applying the same selection procedure to PPs generated in a network trained with List B as target list.

NET2, a new network similar to NET1, was trained either on PPB_A or PPB_B base and tested on both target and control list items. Twelve replications per PP base were run.

Results

As shown in Figure 2, the average error — RMS between the output of NET2 and the tested item — was dramatically smaller for the target than for the control list items [$F(1, 22) = 33{,}160$, $MS_E = 0.000005$, $p < .0001$]. Hence, though drastically selected in order to resemble more the control items than the target items, the PPs generated in NET1 still conveyed efficiently information on the specific (target) items learned by NET1.

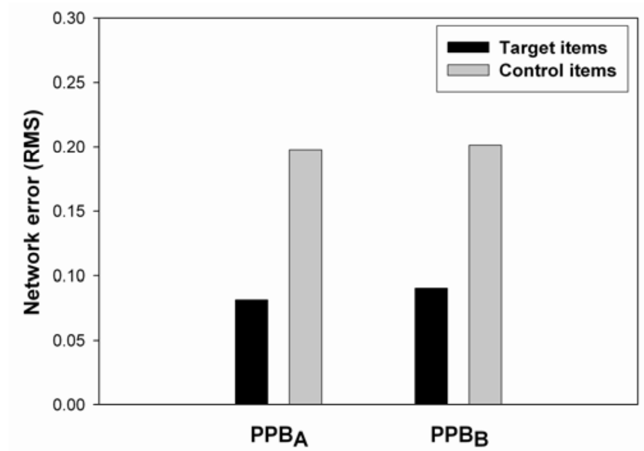

Figure 2: Network performance (RMS error) on target and control items, after training on a PP base (PPB_A or PPB_B, according to a target list counterbalancing).

Behavioral experiment

The general procedure consists in presenting humans incidentally with PPs generated in a GDN auto-associator that had previously been trained on a list of target items. Sensitivity to the information that the selected PPs convey on the target items would be evidenced if humans are shown to have some form of memory for the never seen target items. These items obviously cannot be presented to the participants previous to the test phase, so a task where participants would be instructed to overtly recognize the target items is impossible. Therefore participants' memory will be assessed with an indirect design that allows for a measure of their familiarity with these items.

The experiment is strictly matched with Simulation with respect to the items and the PP lists used. The memory advantage for the never seen target items (over the control items) is tested by comparing perceptual fluency for target and control list items.

After an incidental exposure to the selected PPs used in Simulation, the participants performed a duration judgment task–under time pressure–both on the target and the control list items. Participants were induced to believe that two slightly different presentation times were used and had to classify items' display duration as *short* or *long*. Actually all items had *exactly* the same duration. Participants' subjective impression that a given item "lasts more" is linked to an increased perceptual fluency (Jacoby, 1983; Witherspoon & Allan, 1985), whose real cause is familiarity with that item (Whittlesea, Jacoby & Girard, 1990) — but that participants would attribute to different presentation times. Thus, if humans are sensitive to the distributed information conveyed by PPs, there will be more *long* responses on the target than on the control list items.

Method

Participants Seventy students (mean age = 20.5 years, *SD* = 1.6) participated for course credit.

Stimuli The PPs used in Simulation and the original 106 items were used as stimuli, displayed as 13×13 matrices (260×260 pixels) centered on a black background on a 17" (1024×768 pixels) screen.

Design and procedure Participants first performed an incidental study task: They were to detect a cross that appeared (9 percent of the trials) in a random location on a background made of PPs: PPs were displayed for 400 ms each, with no void in between. Prior to performing this task with the 3000-PP base, and in order to ensure optimal exposure to it, participants underwent a warm-up phase where 500 of the 3000 PPs were displayed.

Then participants engaged in the duration judgment task: They were to classify the display duration of images as *short* or *long*. In order to progressively introduce the test to the participants, a 52-item warm-up phase was designed; in order to prevent interference with test items, only PPs were used during this warm-up phase[1]. During the warm-up, the first 40 trials used two different display durations (200 or 250 ms): The first eight trials were example trials, then participants received feedback on their responses to the remaining 32 trials. The remaining 12 PPs were then presented without feedback and with closer display durations (200 or 230 ms). After this warm-up, participants performed the critical duration judgment task, presented to them as "the same test on a different type of stimuli"; unbeknown to the participants, the presentation time for the 106 items of interest (i.e. target and control items) was in fact always of exactly 200 ms. The inter-stimuli interval was of 1300 ms.

For counterbalancing sake there were two experimental groups: The target list of Group A was List A (and their control list was List B), and the target list of Group B was List B (and their control list was List A).

Results

Responses given by the participants during the first 800 ms are considered. This time limit was chosen in accordance to existing studies that used similar tasks to assess familiarity (e.g. Jacoby, 1991; Ratcliff and McKoon, 1995).

As Figure 3 shows, there were more *long* responses for the target than for the control items [$F(1, 68) = 4.517$, $MS_E = 9.868$, $p = .0372$]. This denotes a stronger familiarity of the participants with the never seen target items (than with the never seen control items), a familiarity grounded in the prior exposure to the (drastically selected) PPs. No other effect was significant–Group effect: $F(1, 68) = 0.370$; interaction: $F(1, 68) = 0.088$.

[1] Both the 500-PP warm-up base used in the incidental study phase and the 52-PP warm-up base used in the duration judgment task comply with the 3 selection rules described in Simulation.

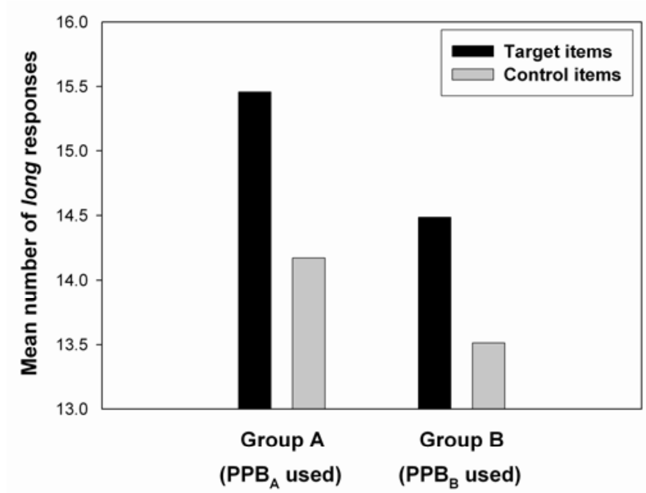

Figure 3: Behavioral results. Effect of previous incidental exposure to PPs on *long* responses (denoting familiarity) to target and control items in a duration judgment task. All stimuli correspond strictly to those used in Simulation (PPB_A was used for Group A and PPB_B for Group B, according to a target list counterbalancing).

Discussion

In this paper it was shown that false memories can be created in humans by exposure to material created in a multi-layered network trained by a gradient descent learning procedure (GDN). The motivation of this research stems from the question of whether humans have the ability to capture distributed information hold in a GDN when presented with samples (reverberated pseudo-patterns, PPs) of the function instantiated by the network. This question is central to a model of human memory that supposes that the final memory (LTM) is similar to a GDN (e.g. McClelland, McNaughton & O'Reilly, 1995) and that memories are dynamically maintained in a dual architecture by means of a memory self-refreshing mechanism based on PPs (Ans & Rousset, 1997; Ans & Rousset, 2000; Ans, Rousset, French & Musca, 2002, 2004; French, 1997).

The results of the behavioral experiment presented here show that the answer is yes. The transport of distributed information has been evidenced in a behavioral experiment using PPs manipulated so as to prevent them from being the exemplars used to train the network or their prototype–under their initial form or as noisy, distorted versions. Even though the selected PPs resembled more the control items than the target items–both at the exemplar and the prototype level–it was shown that the participants presented only with these PPs were more familiar with the target items than with the control items.

What properties of human memory are responsible for these surprising results? To answer this question, we consider the results of Simulation of and other supplementary aspects.

A first element is brought by simulations conducted with two versions of a *matching memory model*. In these models, the similarity of each target and control item to a set of PPs (e.g. PPB_A) is computed and taken as a performance indicator. Two versions of the model were used: One, exemplar-based, computes the similarity between each item and each PP; the other, centroid-based, computes the similarity between each item and the centroid of the set of PPs (i.e. the "mean" PP).

The exemplar-based version of the model can be best described in two steps. First, the similarity $Sim(I_i, PP_j)$ between an item I_i and a pseudo-pattern PP_j is defined as a decreasing function of their RMS:

$$Sim(I_i, PP_j) = \left[1 - RMS\left(I_i, PP_j\right)\right]^{\gamma} \quad (2),$$

where $\gamma \geq 1$ denotes an aggregate parameter.

Then, in order to compare the target and the control lists, a global similarity score is computed for each item as the mean of its similarities to all the PPs. The aggregate parameter γ modulates the contribution of each PP to the global similarity score of the item at hand: For low values of γ all the PPs make a contribution to the global similarity, whereas for high values of γ only PPs (very) close to the item at hand make a significant contribution (thus tending towards a PP-specific matching).

The centroid-based version of the model works in a very similar way, except that the global similarity of each item to PP_M, the centroid of the set of PPs, is used. It is computed in one step:

$$Sim(I_i, PP_M) = \left[1 - RMS\left(I_i, PP_M\right)\right]^{\gamma} \quad (3).$$

Simulations with both versions of the model failed to replicate the behavioral results, whatever the value of parameter γ. This makes it obvious that if human memory functioned as a matching system like those used here, one could not expect the results that the participants exhibit; the nature of the memory system bear crucial consequences on the results.

In order to highlight the crucial role of the learning material and as a further support for the idea that the selected PPs used in Simulation are not mere item distortions, one more set of simulations was conducted. The same network as NET2 in Simulation was used, but with a very different training material. No PPs were created or used at any point in these simulations; instead, mere distortions of the target items were used as training material for NET2.

Distortions of the target items (hereafter *distorted patterns*, DPs) were created by adding a random number lying between 0 and 0.8 (if component's value was 0) or subtracting it (if component's value was 1) to each component of each target item. In order to ensure that DPs and PPs differ only with respect to the absence/presence of information on the function instantiated by NET1 in Simulation, the same three rules used to select the PPs were applied to select among the created DPs.

As expected from the selection rules, when NET2 was trained on mere item distortions, it exhibited a slightly but significantly better performance on the control item list than on the target item list. This result validates the selection rules used and constitutes a clear-cut comparison reference for analyzing the results found in Simulation. When the training material is not made of samples of the function instantiated by a GDN, NET2 not only fails to produce the results found in humans and in the corresponding simulation (Simulation), but produces the opposite pattern of results. This highlights the importance of the learning material: if PPs were just distortions of the target items, the simulation of the behavioral experiment (Simulation) would not have produced the result it did, but the opposite one–just as the simulations that used DPs do.

The behavioral results are simulated when the stimuli used to train the network are samples (i.e. PPs) of the function instantiated by a GDN that has previously been trained on the target stimuli and, importantly, when the memory model used is itself a GDN. After training on PPs, such a system instantiates a function that is very compatible with a good performance on the target items but not on the control items. These considerations seem to point at the conclusion that the participants, after extended exposure to PPs, have captured the function conveyed by the PPs. Under this account, participants' increased familiarity with the target (as compared to the control) items would stem, in a way similar to the GDN trained on PPs in Simulation, from the fact that the target items are obviously compatible with this function.

Because of the thorough controls and the incidental nature of the task used during the acquisition phase, alternative accounts for the behavioral results are scarce: Participants were not asked to attend to the PPs, which served only as a background in the cross-detection task. Thus there was no explicit or implicit request that the participants learn the PPs. Furthermore, all the PPs were obviously part of a single category (that of samples of the function instantiated by a GDN that has previously been trained on the target items), and no categorization was ever required of the participants during the acquisition phase. Two versions of a very simple matching model were considered and they failed to exhibit the result found in the behavioral experiment. There are of course more other theories/models that assume more complex mechanisms (e.g. Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986; Reed, 1972), but they all apply to situations where a categorization task is used during the acquisition phase. Now, this was not at all the case for the experiments presented here.

What are the implications of these results? False memories for never seen items can be created in humans by mere exposure to distributed network attractors in an experimental setting that allows for avoiding an interpretation in terms of prototype or exemplars. The implication of this result is that the actual items are not needed for learning to occur in humans: distributed information coming from a GDN can also induce this learning.

A second implication of the results concerns the nature of memory consolidation. In a GDN architecture, a dynamical and continuous consolidation is required in order to avoid catastrophic forgetting. This constraint of permanent rehearsal pays off, as consolidation is confined to a fully distributed architecture that comes with all the interesting properties of distributed neural networks. As discussed in the introduction, the solution whereby memories are dynamically maintained relies on the use of PPs generated in the GDNs that make up the architecture. This leads naturally to the question whether memory consolidation in the brain is achieved by continuous cortical consolidation in a distributed system similar to a GDN architecture. Of course this question cannot be answered directly. However, if humans could not be shown to be sensitive to the kind of information involved in memory self-refreshing in GDN architectures, one would have good reasons to be prone to give a negative answer. Now, we showed here that false memories can be created in humans through exposure to PPs generated in a GDN. PPs are not just odd, disembodied entities that came up because they are essential to the memory self-refreshing mechanism in a distributed network architecture: Humans are sensitive to and capable of learning from PPs.

Though this paper deals with humans' ability to learn from PPs, the hypothesis of an intra-cortical PP mechanism cannot be tested directly. Because visual modality was chosen to pass the PPs to the human cognitive system and because of the particular experimental design, the results concern phenomena situated at the frontier between memory and perception. Beyond having shown that humans are sensitive to distributed information, a result that supports an original view on the nature of human memory consolidation, our hope is that the experiment presented here also brings new methodological tools for related fields of research.

Acknowledgments

This work was supported in part by a research grant from the European Commission (HPRN-CT-1999-00065) and by the French government (CNRS UMR 5105). We would like to thank Alan Chauvin, Robert French, Dwight Kravitz, James McClelland, David Plaut and Gautam Vallabha for valuable discussions and suggestions on draft versions of this paper.

References

Ans, B., & Rousset, S. (1997). Avoiding catastrophic forgetting by coupling two reverberating neural networks. *Comptes Rendus de l'Académie des Sciences Paris, Life Sciences*, *320*, 989-997.

Ans, B., & Rousset, S. (2000). Neural networks with a self-refreshing memory: Knowledge transfer in sequential learning tasks without catastrophic forgetting. *Connection Science*, *12*, 1-19.

Ans, B., Rousset, S., French, R. M., & Musca, S. C. (2002). Preventing catastrophic interference in multiple-sequence learning using coupled reverberating Elman networks. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 71–76). Mahwah, NJ: Lawrence Erlbaum Associates.

Ans, B., Rousset, S., French, R. M., & Musca, S. C. (2004). Self-refreshing memory in artificial neural networks: learning temporal sequences without catastrophic forgetting. *Connection Science, 16*, 71-99.

Blackmon, J., Byrd, D., Cummins, R., Poirier, P., & Roth, M. (in press). Atomistic learning in non-modular systems. *Philosophical Psychology*.

French, R. M. (1997). Pseudo-recurrent connectionist networks: an approach to the 'sensitivity-stability' dilemma. *Connection Science*, *9*, 353–379.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 21-38.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457.

McCloskey, M., and Cohen, N. J., (1989). Catastrophic interference in connectionist networks: the sequential learning problem. In H. G. Bower (Ed.), *The Psychology of Learning and Motivation, Vol. 24* (pp. 109–165). New York: Academic Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Ratcliff, R. (1990). Connectionist models of recognition and memory: constraints imposed by learning and forgetting functions. *Psychological Review*, *97*, 285–308.

Ratcliff, R., & McKoon, G. (1995). Bias in the priming of object decisions. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 754-767.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382-407.

Robins, A.V. (1995). Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, *7*, 123–146.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*, 716-732.

Witherspoon, D., & Allan, L. G. (1985). The effect of a prior presentation on temporal judgments in a perceptual identification task. *Memory and Cognition*, *13*, 101-111.