Selective memory loss in aphasics: An insight from pseudo-recurrent connectionist networks

In J. Bullinaria, D. Glasspool, & G. Houghton (eds.) *Connectionist Representations*.
Springer, 1997, pp 183-195

Robert M. French
Psychology Department, B32
University of Liège, Liège, Belgium
rfrench@ulg.ac.be

Abstract

McClelland, McNaughton, O'Reilly [15] suggest that the brain's way of overcoming catastrophic interference is by means of the hippocampus-neocortex separation. French [8] has developed a memory model incorporating this separation into distinct areas, using pseudopatterns [23] to transfer information from one area to the other of the memory. This network gradually produces highly compact representations which, while they ensure efficient processing, are also highly susceptible to damage. Internal representations of categories must reflect the variance within the categories. Because the variance within biological categories is, in general, smaller than that in artificial categories and because memory compaction gradually makes all representations proportionately less distributed, representations of low-variance biological categories are likely to be the most adversely affected by random damage to the network. This may help explain the selective memory loss in aphasics of natural categories compared to artificial categories.

1 Introduction

This paper is an attempt to bring together three ideas. The first is that the human brain evolved a particular means of overcoming the problem of catastrophic interference [16, 22] — namely, two separate systems of processing information: the hippocampus in which fast learning takes place and the neocortex where the information learned by the hippocampus is gradually consolidated [15]. A "pseudo-recurrent" connectionist model [8], based on this kind of separation, is described. The second key idea is that connectionist networks reflect the variability of categories in the environment in their internal representational organization. The greater the variability of a category, the greater the variance of the corresponding internal representations. And third, we will extend these ideas to the case of selective memory loss in aphasics. Our ultimate conclusion will be that differences in the variability of various categories may be implicated in preferential memory losses in aphasics of certain categories over others, in particular, the greater loss of natural kinds categories compared to artificial kinds.

In the first part of the paper, we explain the pseudo-recurrent connectionist model [8] that incorporates the idea of two separate storage and processing areas in the human brain in order to prevent catastrophic forgetting. Catastrophic forgetting (also referred to as catastrophic interference) is the tendency of neural networks — whether artificial or natural — to abruptly and completely forget previously learned information upon the learning new input [16, 22]. The vast majority of connectionist networks suffer from this problem because of the highly distributed nature of their internal representations. As French has shown [6], the degree to which networks suffer from catastrophic interference is in part a function of the amount of overlap of internal representations. In other words, the very feature that makes artificial neural networks so powerful and gives them the all-important ability to generalize is the same feature that causes catastrophic interference.

The most generally applied method of avoiding this problem requires a cognitively implausible means of learning new patterns. Whenever the network must learn a set of new patterns, *all* of the patterns it has ever learned in the past must be relearned by the network along with the new ones to be learned. This a far cry from how humans learn new patterns, however. Much of human learning tends to be *sequential*. A particular pattern is learned, then another, and another, and so on. While some of the earlier patterns may be seen again, this is not necessary for them to be retained in memory. As new patterns are learned, forgetting of old, unrepeated patterns occurs gradually as a function of time.

The connectionist architecture presented here is designed, like the brain, to not have to resort to this re-presentation of all of the past patterns it has learned. This means that the network, because it will not catastrophically forget previously learned information like a standard backpropagation network, will be capable of effective sequential learning. This architecture is based on two crucial techniques:

- separating the previously learned internal representations from those that are currently being learned;
- a method of *approximating* the previously learned data (not the original patterns themselves, which the network will not see again) and interleaving these approximations with the new patterns to be learned.

2 The need to separate old and new representations

The most common techniques for reducing catastrophic interference in traditional connectionist architectures have relied on reducing the overlap of representations either by orthogonal recoding of the input patterns [11, 13] or, alternately, by orthogonalizing the network's hidden layer representations [6, 7, 12, 17, 19]. A thorough discussion of these techniques and an analysis of the underlying causes of catastrophic interference can be found in [9, 26]. Pushing the logic of reducing representational overlap to its ultimate conclusion, McClelland, McNaughton, and O'Reilly [15] have argued that the evolution of two separate memory structures, the hippocampus and the neocortex, might have been brain's solution to the problem of new information completely destroying previously learned information. But this stills leaves unanswered a crucial question— namely: *How* does the neocortex store

new information, whether it comes from the hippocampus or elsewhere, without disrupting information already stored there? Their solution involves the very gradual incorporation of the new information into the neocortical structure (i.e., long-term memory). Hippocampal representations very gradually train the neocortex. The problem is that no matter how slowly the hippocampal information is passed to the neocortex, radical forgetting of the old information may still result, *unless a way is found to interleave the already stored neocortical patterns* with the new patterns being learned. This interleaving cannot always use "the rest of the [original] database" [15] of previously learned patterns because many of these patterns will no longer be explicitly available for re-presentation to the network. For example, I have not seen a porcupine in at least ten years, so there has been no re-presentation of this item to my sensory interface during that time and thus no possibility for the corresponding long-term memory concept to be "refreshed" by seeing a real porcupine. Nonetheless, I would have no problem whatsoever recognizing one. There are a great many concepts like this, ones which are not continually refreshed via the environment, but that we still remember without difficulty.

This problem was the reason for the development of the pseudo-recurrent model developed in detail in [8] and exposed briefly here. This architecture has a way to automatically refresh the network without recourse to the original patterns. Instead of the original patterns, internally-produced *approximation*s of these patterns, called pseudopatterns [23], will be used and interleaved with the new patterns to be learned. The architecture proposed will argue for two functionally distinct areas of long-term memory: one, an "early-processing area" in which new information will be initially processed and a second, a "final-storage area," in which information will be consolidated. This model of long-term memory will suggest a natural means of consolidation of information in long-term memory that supports the neurobiologically motivated conclusions of [24].

The ideal way, of course, to solve the problem of catastrophic interference would be to store all previously learned patterns out of harm's way until new input was presented to the system. At that point, all of the previously learned patterns would be taken out of storage, so to speak, and would be mixed with the new patterns. The system would then learn the mixed set of old and new patterns. After the augmented set of patterns had been learned by the network, they would all be put in storage, awaiting the next time new information was presented to the network. There would be no forgetting, catastrophic or otherwise, in this ideal world and new input would have no deleterious effect on the network's ability to generalize, categorize or discriminate.

Unfortunately, this way of learning new data is rarely possible in the real world except in the most artificial situations. It is in essence impossible to explicitly store all, or even a reasonable fraction of previous training exemplars for future learning. I will suggest that *internal approximations* of the original patterns are generated in long-term memory and it is these approximations that, in the absence of the real patterns in the environment, serve to continually reinforce the long-term memory traces of the original patterns. The use of pseudopatterns to improve performance of connectionist networks on catastrophic interference was first proposed by Robins [20] and their plausibility has been further explored in [5, 24].

3 The "pseudo-recurrent" architecture

The architecture discussed in this paper consists of a feedforward backpropagation network that is divided into two parts, one used to help train the other (Figure 1). We will call the left-hand side of the network the "early-processing memory" and the right-hand side the "final-storage memory." It is perhaps easiest to explain how the network works in terms of a specific example.

Suppose that the "final-storage" area contains what the network has learned up to the present time. The network is then asked to sequentially learn 20 new patterns, P_1, P_2, ... P_{20}. Each of these patterns, P_i, consists of an input and an output ("teacher") association: (I_i, T_i). By sequentially learning these patterns we mean that each individual pattern must be learned to criterion before the system can begin to learn the subsequent pattern. To learn pattern P_1, its input I_1 is presented to the network. Activation flows through both parts of the network, but the output from the final-storage part is prevented from reaching the teacher nodes by the "real" teacher T_1. In other words, the teacher pattern T_1 fills the teacher nodes. The early-processing network then adjusts its weights with the standard backpropagation algorithm using as the error signal the difference between T_1 and the output O_1 of the early-processing network. Crucially, however, the early-processing network does not only learn the pattern P_1. Internally created *pseudopatterns*, reflecting the contents of final-storage, are also generated by the final-storage memory and will be learned by the early-processing memory along with P_1.

Pseudopatterns are generated by final-storage and learned by the early-processing memory as follows. A random input pattern, i_1, is presented to the input nodes of the system. This input produces an output, o_1, at the output layer of the early-processing memory and also produces an output, t_1, on the teacher nodes of the final-storage memory. This input-output pair (i_1, t_1) defines a pseudopattern, ψ_1, that reflects the contents of the final-storage memory. The difference between t_1 and o_1 determines the error signal for changing the weights in the early-processing memory. Similarly, the other random inputs, i_2, i_3, . . . i_n, produce pseudopatterns, ψ_2, ψ_3, . . . ψ_n that are also be learned by the early-processing memory. Once the weight changes have been made for the first epoch for the set of patterns {P_1, ψ_1, ψ_2, . . . ψ_n}, the early-processing memory cycles through this set of patterns again and again until it has learned them all to criterion. By learning the pattern P_1 the early-processing memory is learning the new information presented to it; by learning the pseudopatterns $\psi_{1, \ldots,} \psi_n$, the early-processing memory is, in addition, learning an approximation of the information previously stored in final storage. Obviously, the more pseudopatterns that are generated, the more accurately they will reflect the contents of final storage. Once learning in the early-processing network has converged for P_1, ψ_1, ψ_2, . . . ψ_n, the early-processing weights then replace the final-storage weights. In other words, the early-processing memory *becomes* the final storage memory and the network is ready to learn the next pattern, P_2. (Note that this weight-copying strategy is certainly not biologically plausible. However, it has been shown [8] that

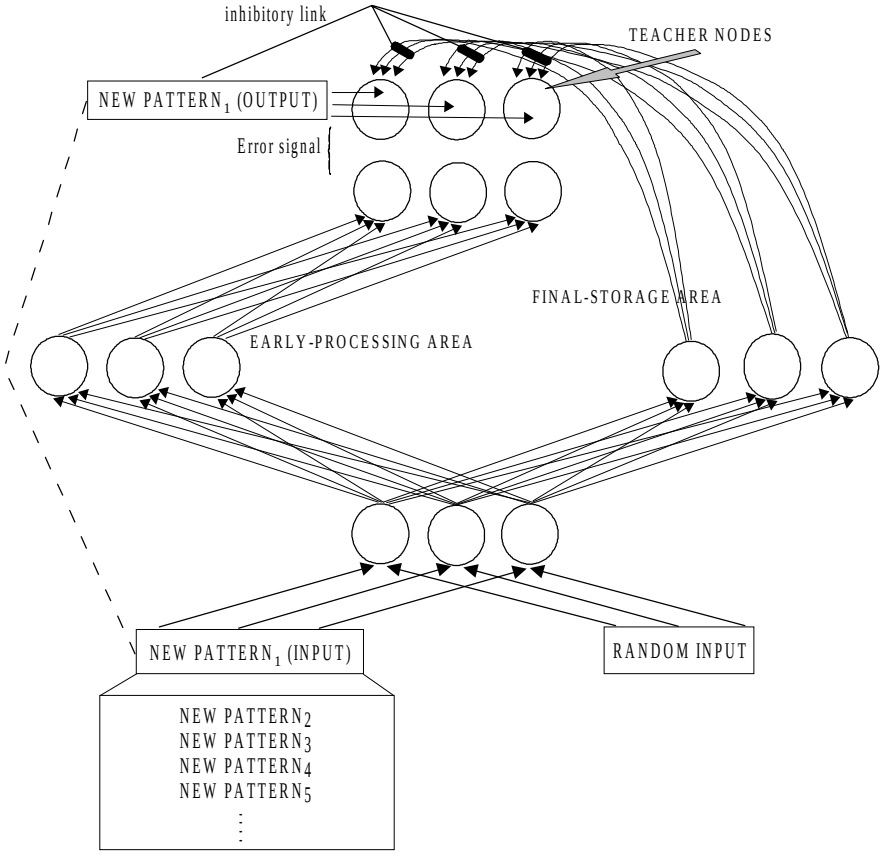

Figure 1. The pseudo-recurrent network architecture

information transfer can also be effectively done from early-processing to final-storage by means of the above type of pseudo-pattern transfer.)

The essence of this technique is to interleave new information to be learned with pseudopatterns that reflect the contents of final-storage. Thus, rather than interleaving the real, originally learned patterns with the new input coming to the early-processing memory, we do the next best thing — namely, we interleave pseudopatterns that are *approximations* of the previously stored patterns. Once the new pattern and the pseudopatterns are learned in the early-processing area, the weights from the early-processing network are copied to the corresponding weights in the final-storage network (or, more plausibly, the early-processing area trains the final-storage area using its own set of pseudopatterns).

The model is called "pseudo-recurrent" not only because of the recurrent nature of the training of the early-processing memory by the final-storage memory — approximations of previously learned information is continually fed back into the early-processing area from final-storage —, but also as a means of acknowledging

the all-important mechanism of information transfer from final-storage to early-processing storage — namely, pseudopatterns.

3.1 Testing the pseudo-recurrent network

This type of network has been extensively tested and it has been shown not to suffer from catastrophic interference [8]. As a result, it is able to do sequential learning in a cognitively plausible manner. In Figure 3 the performance difference of the pseudo-recurrent network on a sequential learning task is compared with standard backpropagation. We will briefly describe this experiment designed to illustrate the pseudo-recurrent network's crucial ability to do sequential learning.

In these tests on the pseudo-recurrent network, we will show that it does two things that will allow us to use this model to provide a possible insight into selective memory losses in aphasics — namely:

- Sequential learning: The network does not forget catastrophically and, as a result, can learn sequentially. It is, therefore, a more cognitively plausible memory model than standard backpropagation networks, both in terms of its architecture and its performance.
- Emergence of compact representations: Over time, the network automatically develops "compact" internal representations in the final-storage area. This gradual representational compaction has numerous desirable effects, especially in terms of resource utilization, but has a major negative consequence: compact representations are more easily damaged by any damage to the network.

3.1.1 Sequential Learning

To test the network on sequential learning, we used the 1984 U.S. Congressional Voting Records database from the University of California at Irvine repository of machine learning databases [18]. Twenty members of Congress (10 Republicans, 10 Democrats, defined by their voting record on 16 issues) were chosen randomly from the database. Each of the 20 patterns presented to the network therefore consisted of 16 binary inputs and a single binary output. The 20 patterns were presented *sequentially* to the network. In other words, a new pattern was presented only after the previous one had been learned to criterion (i.e., the output on the single output node was within 0.2 of the desired output). After all twenty patterns had been sequentially learned by the network, a test was made of the percentage of these patterns that the network still correctly remembered. A pattern was considered to have been exactly remembered by the network if, when it was presented to the network, the output that was produced on the single output node remained within 0.2 of the desired output for that pattern. After sequentially learning the 20 patterns, Figure 2 shows that a standard backpropagation network can exactly remember about 40% of the them. This figure climbs rapidly to 65% when 5 pseudopatterns from final-storage are added along with each new pattern to be learned. With 50 pseudopatterns added to each new item to be learned, exact recognition of all of patterns climbs to 85%.

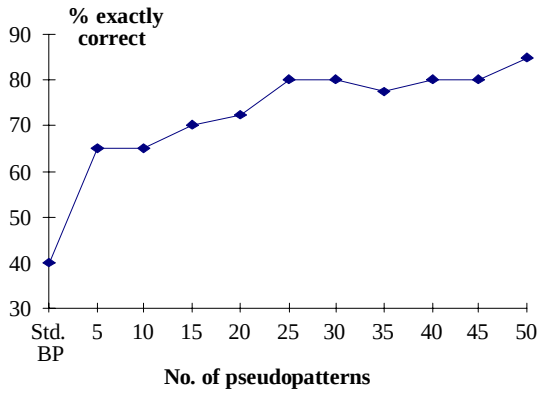

Figure 2. Percentage of all data exactly recalled by the network after serial presentation of all 20 patterns (median data).

After the network learned the twenty items sequentially, each item was tested to see how well the network remembered that individual item. The hypothesis was that forgetting would be more gradual when pseudopatterns were used compared to standard backpropagation. Figure 3 shows that this is indeed the case. Of particular importance is the difference in amount of forgetting between the final item learned and the penultimate one for both standard backpropagation and the 25-pseudopattern network. Further, as can be seen in the figure, the standard backpropagation network is, on average, significantly above the 0.2 convergence criterion for *all* of the previously learned items, whereas the pseudo-recurrent network is at or below criterion for the last eight items learned (items 13-20) and within 0.05 of the criterion for items 7-12.

Clearly, forgetting is taking place more gradually in the pseudo-recurrent network than in the backpropagation network, where *none* of the 19 previously learned items are below criterion after the 20th item has been learned (Figure. 3).

This experiment shows that the forgetting curves for this type of network are considerably more gradual than with standard backpropagation. This experiment also points out the importance of interleaving approximations of the already-learned patterns with the new patterns to be learned.

3.1.2 The emergence of compact representations

During sequential learning, information is continually passed back and forth between the two memory areas by means of pseudopatterns. (In the version of the model described above pseudopatterns are used only to transfer information from final-storage to the early-processing area. Weight copying is used in the other direction. This constraint has been successfully relaxed and pseudopatterns have been used to pass information in both directions. A detailed discussion of these results can be found in [8].) It turns out that one unanticipated result of this use of pseudopatterns is the "compaction" of the representations that develop in final storage. This has numerous advantages, among them, a decrease in the number of

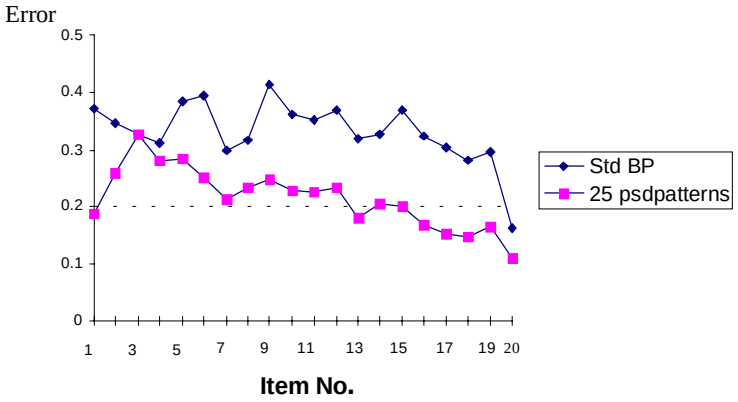

Figure 3. Amount of error for each of the 20 items learned sequentially after the final item has been learned to criterion (in this case, 0.2).

resources required to activate any given concept, a decrease in the amount of overlap in final storage, etc. On the other hand, compact representations are more likely to be destroyed if the network is lesioned.

This suggests an interesting possibility for the human brain. It has been shown that there is continual interaction between the hippocampus and the neocortex and that this interaction is almost certainly involved in long-term memory consolidation. If this interaction is indeed mediated by pseudopatterns, as suggested by Robins [24], then it would not be unreasonable to think that the representational compaction observed in the pseudo-recurrent model might also occur in the brain. Compact representations would, presumably, allow for more efficient processing of incoming stimuli because of their reduced demand on system resources (i.e., less activation is required to fully activate a compact representation, fewer connections are involved, etc.). On the other hand, the more compact these representations, the more difficult it would be to make category distinctions (see [7] for a discussion of this problem). This would also lead to the category brittleness that occurs as people grow older, a phenomenon whose description dates at least to James [10], and results in a loss of the capability of "assimilating impressions in any but old ways." [10] Finally, highly compact representations would presumably be more vulnerable to severe disruption than highly distributed representations. This, too, would seem to be consistent with selective memory loss with aging.

The data used to test the network were obtained from measurements of the original Cat and Dog pictures used by Eimas, Quinn, and Cowan [4, 21] and by Mareschal and French [14] to study infant categorization. They included 18 dogs and 18 cats classified according to head length, head width, eye separation, ear separation, ear length, nose length, nose width, leg length vertical extent, and horizontal extent.

A 10-30-10 autoassociator was used in learning these two categories of animals. We compared the hidden unit representations in networks that had sequentially learned 6 pairs of cats (i.e., a total of 12 cats) using differing numbers of pseudopatterns. The network was then tested on its ability to correctly autoassociate

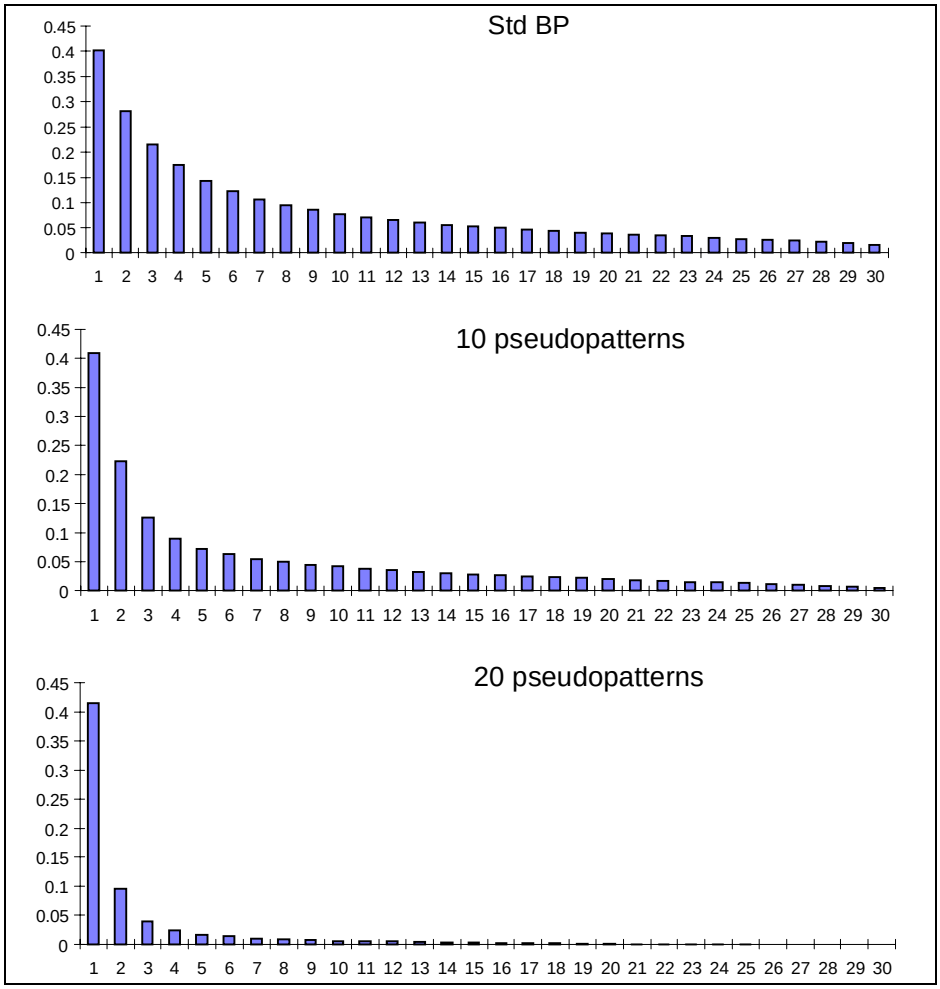

Figure 4. Pseudopatterns gradually produce highly "compact" representations for the categories in final storage (in this case, for the category "cat"). The greater the number of pseudopatterns, the more compact the representations becomes.

the previously learned cats. As the number of pseudopatterns increases, the network's internal representation of the concept "cat" undergoes compaction (Figure 4).

It is the continual interaction between the two processing areas of the pseudo-recurrent memory that gives rise to these more compact (i.e., less distributed) internal representations. These are the same types of representations that have been shown in other work to reduce catastrophic interference [6, 7]. Unlike this earlier work in which explicit algorithms produced this type of representation to reduce catastrophic interference, in this memory model — and conceivably in the brain as well — they emerge naturally from the mechanisms underlying the model.

4 Selective memory loss in aphasics

Natural kinds are categories like “cat”, “bird”, “horse”, etc., while artificial kinds are categories like “chair”, “house”, “clothes”, etc. It has been repeatedly shown [1, 25, 28] that there can be selective memory loss in certain aphasics for natural-kinds compared to artificial-kinds categories. In contrast to explanations that rely on the form/function distinctions in their attempts to explain the observed selective anomia (for example, [3]), I will suggest that this selective memory loss is due, at least in part, to the considerable difference in the average variability within most biological and artificial kinds. This difference, combined with the phenomenon of gradual compaction of representations as they are consolidated in final-storage — making them increasingly susceptible to damage — will provide a simple, if undoubtedly partial and speculative, account for this type of aphasia.

If two real-world categories that have very different variance are stored in a network — connectionist or human — this difference in variance must be reflected in a difference in the variance of the internal representations of the two categories. The greater the variance in the real-world category, the greater the variance in the internal representation of that category, where the variance of an internal representation is determined by the “spread” of the distribution of hidden-unit activation pattern corresponding to a representation when it is activated. The more spread out the distribution, the greater the variance. Consider, for example, the artificial-kind category “house” and the natural-kind category “cat.” The former has greater variance than the latter. The folk observation that “If you’ve seen one cat, you’ve seen ’em all” translates more rigorously into a statement about the lack of variability within the category of cats. On the other hand, the same could never be said about all houses, which certainly *do not* all look the same. Some are brick, some wood, some tall, some wide, some are made of logs, some of stone, some even of cloth (a “teepee,” for instance) or animal hides, etc. This greater category variance will be reflected in a greater amount of variance in the internal representations for each category [14]. So, for example, we might have internal representations for “house” and “cat” similar to those in Figure 5.

But if, as it has been suggested [24], neural pseudopatterns are used to consolidate the long-term memory trace, the representations shown in Figure 5 will gradually become more and more compact, giving rise to representations like those in Figure 6. And, since compaction occurs in a uniform manner, the representations that were more highly distributed initially will remain so as compaction progresses.

The problem, though, is that, while these more compact representations will certainly be more efficient in terms of overall processing demands, they are also more easily disrupted. If we randomly remove, say, four nodes from the hidden-layer of the above network, the chances of destroying the internal representation for “cat” will be far greater than for destroying “house.” Of course, one prediction of the pseudo-recurrent model is that, on rare occasions, we should also see an anomia for an artificial-kind category, but that this should be far less frequent than for natural-kind categories. Anomia of this kind has been observed, but is, indeed,

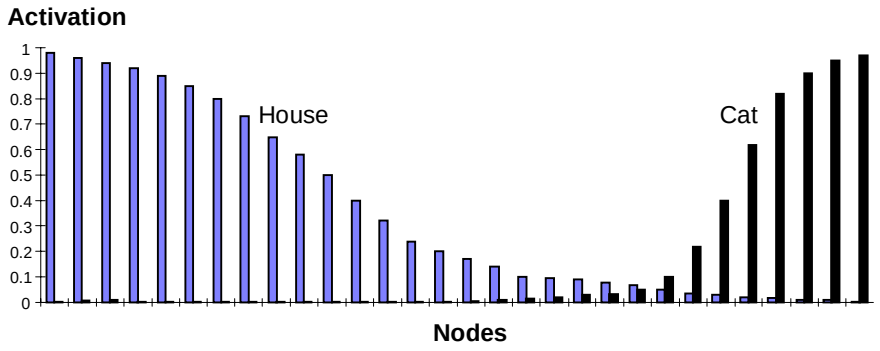

Figure 5. The internal representations of the categories "house" and "cat" before the gradual effect of pseudopattern transfer transforms them into the more compact representations shown in Figure 6.

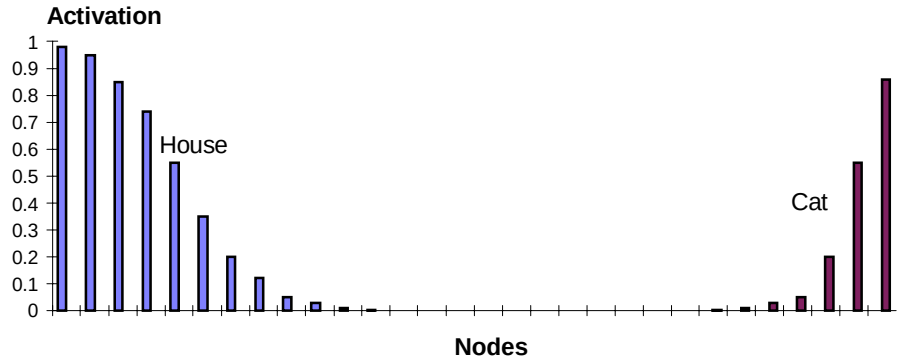

Figure 6. The representations for "house" and "cat" after undergoing gradual compaction due to the effect of pseudopattern transfer.

exceedingly rare. While there is a growing body of research using connectionist models to understand category-specific deficits [2, 3, 20, 27, etc.], the above model is unique in suggesting that this phenomenon might, at least in part, be related to gradual representational compaction in long-term memory and subsequent damage to these compact representations.

5 Conclusion

This paper proposes a "pseudo-recurrent" connectionist model of long-term memory in which an "early-processing" memory and a permanent, "final-storage" memory continually exchange information by means of pseudopatterns [8]. Some form of neural pseudopatterns arguably play a role in long-term memory consolidation [24]. However, one of the consequences of pseudopattern transfer is a gradual compaction of the representations stored in long-term memory. I argue that this compaction,

along with the fact that the internal representation of a category must reflect the amount of variance of that category in the environment, may contribute to the observed selective memory losses in aphasics. In particular, the greater variance, on average, of artificial kinds compared to natural kinds would predict greater losses of the latter kind of category, which corresponds to aphasia data. This is certainly not the whole story on selective memory loss in aphasics, but this paper suggests that it could be an important part of it.

Acknowledgments

This work was supported in part by Belgian National Science Grant PAI P4/19. The author also wishes to thank Jean-Pierre Thibaut for his significant contribution to the ideas of this paper.

References

1. Brandt, R. (1988). Category-specific deficits in aphasia. *Aphasiology, 2*(3-4) 237-240.
2. Devlin, J., Gommerman, L., Anderson, E., Seidenberg, M. (1997). Category specific deficits in focal and widespread brain damage: A computational account. (Submitted to *Journal of Cognitive Neuroscience*).
3. Durrant-Peatfield, M., Tyler, L., Moss, H. & Levy, J. (1997). The distinctiveness of form and function in category structure: A connectionist model. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, Mahwah, NJ:LEA.
4. Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants, *Journal of Experimental Child Psychology, 58*, 418-431.
5. Frean, M. and Robins, A. (1996). Catastrophic forgetting: A review and an analysis of the pseudorehearsal solution. (under review).
6. French, R. M. (1992). Semi-distributed representations and catastrophic forgetting in connectionist networks. *Connection Science, 4*, 365-377.
7. French, R. M. (1994). Dynamically constraining connectionist networks to produce orthogonal, distributed representations to reduce catastrophic interference. In *Proceedings of the 16th Annual Cognitive Science Society Conference*. Hillsdale, NJ:LEA, 335-340.
8. French, R. M. (1997). Pseudo-recurrent connectionist networks: An approach to the "sensitivity–stability" dilemma. *Connection Science, 9*(4) (in press).
9. Hetherington, P. A. (1991). The sequential learning problem in connectionist networks. Unpublished Master's Thesis, Psychology Department, McGill University, Montreal.
10. James, W. (1890). *Psychology, The Briefer course*. New York: Holt.

11. Kortge, C. (1990). Episodic Memory in Connectionist Networks, In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: LEA, 764-771.
12. Kruschke, J. K. (1993). Human Category Learning: Implications for Backpropagation Models. *Connection Science*, *5*(1), 1993.
13. Lewandowsky, S. & Shu-Chen Li (1993). Catastrophic Interference in Neural Networks: Causes, Solutions, and Data. In *New Perspectives on interference and inhibition in cognition* F.N. Dempster & C. Brainerd (eds.). New York, NY: Academic Press.
14. Mareschal, D. & French, R. (1997). A connectionist account of interference effects in early infant memory and categorization. In *Proceedings of the 19th Annual Cognitive Science Society Conference*, New Jersey: LEA (in press).
15. McClelland, J., McNaughton, B., & O'Reilly, R. (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.
16. McCloskey, M. & Cohen, N. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *The Psychology of Learning and Motivation, 24*, 109-165.
17. McRae, K. & Hetherington, P. (1993) Catastrophic interference is eliminated in pretrained networks. In *Proceedings of the 15h Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: LEA. 723-728.
18. Murphy, P. & Aha, D. (1992). UCI repository of machine learning databases. Maintained at the Dept. of Information and Computer Science, UC Irvine, CA.
19. Murre, J. (1992). The effects of pattern presentation on interference in backpropagation networks. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum. 54-59.
20. Plaut, D.C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropschology, 17*, 291–326.
21. Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants, *Perception, 22*, 463-475.
22. Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions, *Psychological Review, 97*, 285-308.
23. Robins, A. (1995). Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, *7*(2), 123–146.
24. Robins. A. (1996). Consolidation in Neural Networks and in the Sleeping Brain. (under review).
25. Sartori, G., Miozzo, M., & Job, R. (1993). Category-specific form-knowledge deficit in a patient with Herpes Simplex Virus Encephalitis. *The Quarterly Journal of Experimental Psychology, 46A*(3) 489-504.
26. Sharkey, N. & Sharkey, A., (1995). An analysis of catastrophic interference. *Connection Science*, *7*(3-4), 301–329.

27. Small, S. (1994) Connectionist networks and language disorders. *Journal of Communication Disorders*, *27*, 305-323.
28. Temple, C. (1986). Anomia for animals in a child. *Brain*, *106*, 1225-1242.