

Modeling time perception in rats: Evidence for catastrophic interference in animal learning

Robert M. French and André Ferrara

Department of Psychology (B32)

Univeristy of Liège, Liège, Belgium

email: rfrench@lg.ac.be; a.ferrara@ulg.ac.be

Abstract

For all intents and purposes, catastrophic interference, the sudden and complete forgetting of previously stored information upon learning new information, does not exist in healthy adult humans. But does it exist other animals? In light of recent research done by McClelland, McNaughton, & O'Reilly (1995) and McClelland & Goddard (1996) on the role of the hippocampal-neocortical interaction in alleviating catastrophic interference, it is of particular interest to ascertain whether catastrophic interference occurs in non-human higher animals, especially in those animals with a hippocampus and a neocortex, such as the rat. In this paper, we describe experimental evidence to support our claim that this type of radical forgetting does, in fact, exist for certain types of learning in some higher animals, specifically, in the rat's learning of time-durations. We develop a connectionist model that could provide an insight into how the rat might be encoding time-duration information.

Introduction

Catastrophic forgetting is a well-known problem in connectionist modeling in which the learning of new information causes the sudden and complete disappearance of previously-stored information. (For a review, see French, 1999). The severity of this problem first came to light at the end of the 1980's (McCloskey & Cohen, 1989; Ratcliff, 1990) and has been the subject of on-going research since. In healthy adult humans, however, there is no evidence of catastrophic forgetting. Humans, it seems, learn — and unlearn — gradually (Barnes & Underwood, 1959). But does catastrophic forgetting exist in non-human animals? We will present experimental results to show that catastrophic interference *does* seem to exist for certain types of learning in higher animals, specifically, in the rat's learning of time durations. We then develop a connectionist model that could provide an insight into how the rat might be encoding time information.

Catastrophic interference in connectionist networks is due, at least in part, to the overlapping nature of the network's distributed internal representations. The smaller this overlap, the less the amount of catastrophic interference (French, 1991). Various algorithms have been proposed to reduce internal representational overlap in order to decrease catastrophic interference. These algorithms generally rely on explicitly manipulating the hidden-layer representations (e.g., French, 1991, 1994; Murre, 1992; Krushke, 1992) or on orthogonalizing the input representations with the expectation that this will decrease internal representational overlap (e.g. Lewandowsky, 1991; Lewandowsky & Su-

Chen, 1993). These techniques do, in fact, produce significantly reduced catastrophic interference.

The idea of reducing the overlap of internal representations was carried to its logical conclusion in McClelland, McNaughton, and O'Reilly (1995) and McClelland & Goddard (1996). They suggested that the reason humans show no signs of catastrophic interference was because of our two complementary learning systems: the hippocampus and the neocortex. Their idea was that new information is initially learned in the hippocampus, where it cannot not adversely affect information that has been previously consolidated in the neocortex. Once the new information was learned in the hippocampus, it was transferred very gradually to neocortical long-term storage. In this way, previously learned information could be kept "out of the way" of newly arriving information. This, they claimed, was how the brain overcame catastrophic forgetting.

Since their theory is based on the key notion of a dual hippocampal/neocortical mechanism of early-processing and subsequent storage, we wondered whether there was evidence of catastrophic forgetting in animals that, like humans, had both a neocortex and a hippocampus. We believe we have discovered a likely candidate for catastrophic forgetting in animals having both hippocampus and neocortex — namely, time-duration learning (and forgetting) in the rat. Our results suggest that the hippocampal/neocortical loop may not be involved in time learning in the rat, or if it is, it may not function as it does for types of learning that are not subject to catastrophic interference. We show that a simple single-store connectionist network can provide a relatively good model this type of time-duration learning. We will conclude by suggesting that the greater amount of overlap of the network's internal representations of time-durations acquired during sequential learning compared to concurrent learning may be why catastrophic forgetting occurs after sequential learning and is absent in the case of concurrent learning. We suggest that there might be a counterpart to these differences in the internal time-duration representations in the rat.

Perception of time-duration in the rat

In their natural environment, animals are, of necessity, good at predicting significant events such as periodical food availability. This seems to be a clear indication of an ability to represent time. Laboratory researchers studying time-learning in animals have developed a number of techniques to study timing processes. One of these is the *peak procedure* (Catania, 1970) in which rats learn to press a

lever to receive food after a certain fixed duration. Each trial in this procedure begins with the simultaneous onset of a sound stimulus and the insertion of a lever into the Skinner box. There are two different types randomly mixed of trials. During *reinforced (or "food") trials*, only the first lever press after the critical duration is rewarded with a little food pellet. Immediately following the reinforced lever-press, the sound is switched off and the lever is withdrawn from the box. *Test trials* begin exactly like reinforced trials, but the animal receives no reward. Test trials are necessary because in the reinforced trials, the animal's lever-pressing stops as soon as the food pellet drops into its box. These trials end independently of the lever presses made by the animal and typically last at least twice as long the duration reinforced in the rewarded trials. The number of lever presses are recorded for each one-second interval and averaged across test trials. This produces the characteristic bell-shaped response-rate function observed with this procedure (Fig. 2). The moment of maximum lever-pressing is called the *peak time* and reflects the moment of maximal food expectation by the rat.

The observation of steady-state behavior following training has long been used to understand the mechanisms underlying timing abilities (Roberts, 1981; Gibbon, Church, & Meck, 1984). More recently, it has also been used to study the acquisition of a new temporal representation (Meck, Komeily-Zadeh, & Church, 1984; Ferrara, 1999).

Time perception experiment

Sequential time-duration learning

A group of 15 rats was used in a learning experiment that studied behavioral adaptations to changes in rewarded time-durations. How rats adapt their behavior when learning time-durations is assumed to reflect properties of its temporal representation. The present experiment was divided in several phases. In the first phase the animal was trained on a 40-sec. duration (14 days, with 90 trials per day of which 85% were reinforced trails and 15 % test trials). In the next phase, the animal learned an 8-sec. duration (10 days, 90 trials per day with the same proportions of fixed and test intervals).

The moment when the animals were switched from 40-sec. learning to 8-sec. learning is referred to as Transition 40-8 (1) and can be seen in the upper left graph of Figure 2. After the 8-sec. duration had been learned, the reward was switched back to the original 40-sec. duration (Transition 8-40 (2) in the lower left graph of Figure 2).

The rate of lever-pressing was recorded during three different sessions: just before a transition, the two sessions immediately following a transition. The final session of Phases 1, 2, 5 and 6 (i.e., the last session before switching to a new time-duration) is referred to as a *reference session* (indicated by a (1) in Fig. 1). The peak time of the reference session reflects the moment of maximum expectancy for

receiving a food pellet. The *transition session* (indicated by a (2) in Fig. 1) refers to the first session after switching the animal to a new time-duration. It is interesting to notice how the peak time shifts during this session compared to the reference session, since, presumably, any changes in peak time reflect modifications of the animal's internal representation of the reinforced duration. The *transition+1 session* (indicated by a (3) in Fig. 1) refers to the second session after the transition. Learning a new time-duration can be described in terms of a moving peak time. In other words, during Transition 40-8 (1), the peak time shifts from the previously learned 40-sec. duration and stabilizes around the new 8-sec. duration which is being reinforced (Fig. 2). Similarly, in Transition 8-40 (2), the peak time will gradually move from 8 to 40 seconds. It has been shown elsewhere (Lejeune, Ferrara, Simons & Wearden, 1997) that, even if the reinforcement is repeatedly switched back and forth from one time-duration to the other, there is no real improvement in the speed of the peak time adaptation. In fact, re-learning a previously learned time-duration requires as much time as learning an entirely new one. This suggests that at each transition, the previous representation of the reinforced time is "overwritten" as the new one is built. In other words, *there is no evidence of any memory savings from the previously learned time-durations*. One reasonable interpretation of this result is that new time-duration learning completely (catastrophically) wiped out the originally learned time-duration.

Concurrent learning

In the next phase (Phase 4), the rats were *concurrently* trained for 25 consecutive sessions on both 8- and 40-sec. durations. This was done by means of a random mixture of 42.5% 8-sec. trials, 42.5% 40-sec. trials and 15% test trials (each session still contains 90 trials). At the end of this learning period, the response curves (not shown here) became clearly bimodal, with a first peak located around 8 seconds and a second peak located around 40 seconds. This means that during concurrent learning of the two time-durations, the animal had developed a representation for both durations.

Phases 5, 6 and 7 were identical to phases 1-3 (except for the number of sessions). Transition 40-8 (3) and transition 8-40 (4) show the curves for lever-pressing rates for the reference, transition and transition+1 sessions after the animals have learned 40- and 8-sec. durations concurrently as above.

Unlike the previous case of sequential learning in which there was no savings of prior learning, the animal, having now learned the two durations concurrently, can rapidly shift from the 40-sec. duration back to the 8-sec. duration. In this case, while there is still a small amount of forgetting, *there is no catastrophic forgetting of the originally learned 8-sec. duration*. This seems to imply that time-

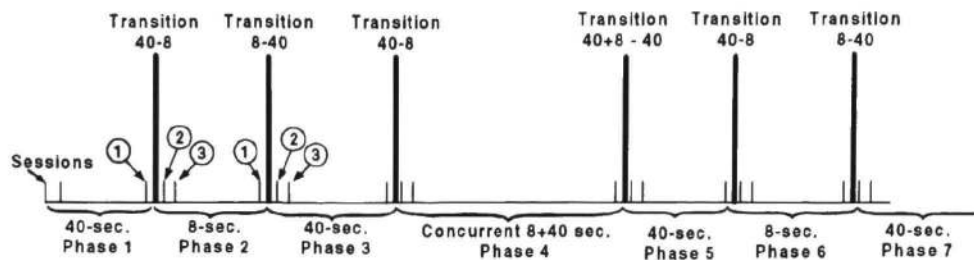


Figure 1. Lever-pressing rates were recorded during the reference session (1) at the end of one learning period and during the transition (2) and transition+1(3) sessions, i.e., the first and second sessions after switching to a new learning period.

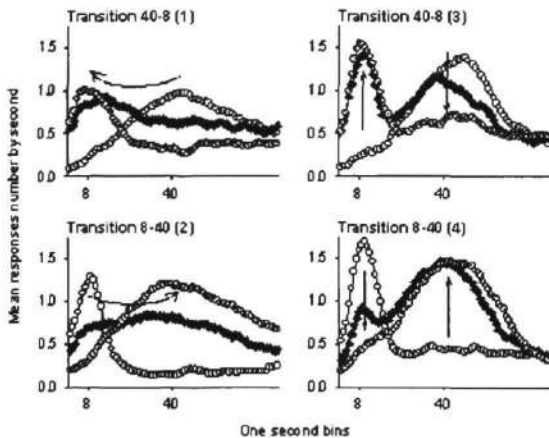


Figure 2. Average rates of lever-presses/second for the reference (white), transition (black) and transition+1 (gray) sessions.

duration representations developed during concurrent learning are significantly different than those developed during sequential learning.

Simulation

To simulate these results, we used an 11-18-2 backpropagation network with binary input coding. The learning rate was set at 0.1, momentum 0.9. The learning criterion was set at 0.001 for all output nodes. An 11-unit input layer was used because this corresponds to the oscillation periods used in the Church and Broadbent (1990) model. Eleven oscillation periods were chosen by these authors because this is enough to represent the full range of relevant short-term time-durations experienced by the animal. The oscillation periods used in the following simulation were as follows: 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8, 25.6, 51.2, 102.4, 204.8 seconds.

Each time duration was translated into a binary pattern of these 11 oscillators, with each oscillator either being activated (= 1) or not (= 0). On each training session the input patterns were modified by the addition of gaussian noise. The noise added to each oscillator signal was proportional to its mean oscillation period (i.e., the larger the oscillator period, the larger the normal curve around its mean). Target outputs of 00 indicated no reinforcement; whereas 11 indicated a positive reinforcement.

Each "session" consisted of a grouped presentation of 20 patterns, P_i , learned for 16 epochs (or until the error for all patterns was below 0.001). Each pattern consisted of an input, which was an encoding for a particular time-duration and a desired output, corresponding to whether or not that particular time-duration was reinforced. For example, an "8-sec. only" session is:

P_1 : Input: 2-sec.	Output: 00
P_2 : Input: 4-sec.	Output: 00
P_3 : Input: 6-sec.	Output: 00
P_4 : Input: 8-sec.	Output: 11 ("reinforced")
...	
P_{20} : Input: 40-sec.	Output: 00

There were six distinct phases for sequential-learning, five for concurrent-learning simulations (Figs. 3, 4).

Phase 1 and 2 were identical for both simulations. The first consisted of 20 "8-sec. only" sessions (i.e., only the 8-sec. duration was reinforced); the second by 20 "40-sec. only" sessions (i.e., only the 40-sec. duration was reinforced).

The next phases were critically different for the two simulations. For sequential learning, Phases 3 consisted forty "8-sec. only" sessions followed by Phase 4 which had forty "40-sec. only" sessions. For the concurrent learning simulation, however, these two phases were combined into a single 80-session phase during which both 8-sec. and 40-sec. durations were reinforced.

The test phases (i.e., Phases 5 and 6 for sequential learning and Phases 4 and 5 for concurrent-learning) in both simulations were identical and consisted of 40 sessions. The first of the two test phases consisted of eight "8-sec. only" sessions. This was long enough for the previously learned 40-sec. duration to be "forgotten" by the network, after having learned it in Phase 4 in the sequential simulation and, along with the 8-sec. duration, in Phase 3 of the concurrent simulation. In other words, "unlearning" the 40-sec. duration, whether in the sequential simulation or the concurrent simulation, meant that when 40 seconds was input to the network, it correctly responded with 00. This took eight sessions. The final 32 sessions of this phase were "40-sec. only" sessions. The critical observation was how quickly the network recovered its knowledge of the 40-sec. time duration.

All results reported in this paper were averaged over 60 independent runs of the network..

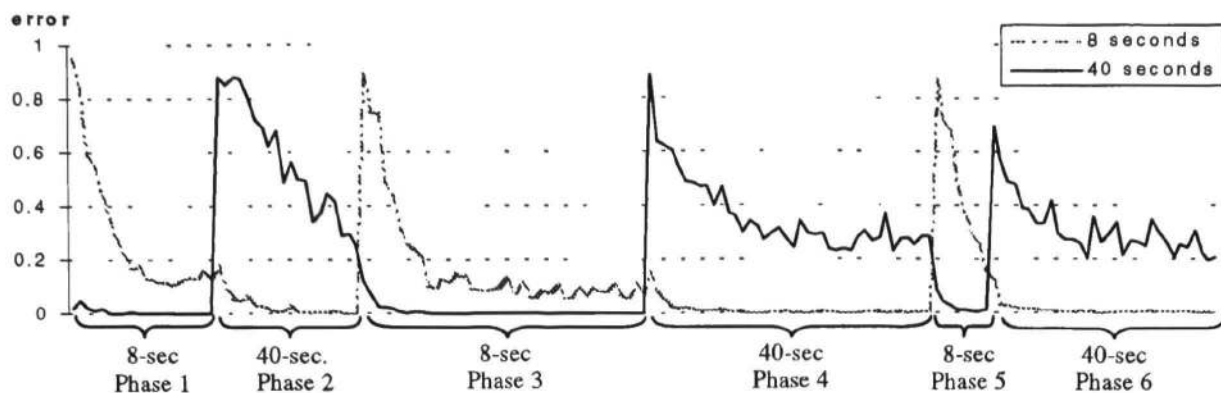


Figure 3. Phases of 8-sec. and 40-sec. learning when these durations are learned sequentially.

Results

Sequential learning

We began by running the network in “sequential mode” (Fig. 3). At the beginning of Phase 3, when the network returns to learning the 8-sec. duration, its error curve is virtually identical to the first time it ever encountered the 8-sec. duration in Phase 1. In other words, learning the 40-sec. duration in Phase 2 seems to have completely erased any memory trace of the initial 8-sec. learning. There is no evidence of any “savings” of the initial 8-sec. learning. (Ebbinghaus, 1887; see Hetherington & Seidenberg, 1989, for a discussion of this as a measure of catastrophic forgetting).

In short, the network is responding to learning new time durations in much the same way as the rat: learning a new duration seems to completely erase (or overwrite) the memory trace of the prior time-duration learning.

Concurrent learning

In the second simulation we explore concurrent time-duration learning and find that it produces considerably different results compared to learning time-durations sequentially (Fig. 4). Phases 1 and 2 in the concurrent-learning simulation are identical to Phases 1 and 2 in the sequential-learning simulation. However, now in Phase 3 the network learned both 8-sec. and 40-sec. durations concurrently for 80 sessions (instead of 40 8-sec. sessions

followed by 40 40-sec. sessions as in the previous simulation). After this concurrent learning phase, the network is tested as in the previous simulation. In other words, in Phase 4, the network is trained for eight sessions on 8-sec. durations only. As before, this was long enough for its performance on 40-sec. durations to return to a zero-error baseline. Reinforcement was then switched back to a 40 seconds (Phase 5, Fig. 4)

Now let us compare the test phases in the two simulations. In both simulations, the eight sessions of 8-sec. learning only, allowed the network to return to the unreinforced baseline for 40-sec. learning (see the 40-sec. error rate in Phase 5 in the sequential-learning simulation, and in Phase 4 in the concurrent-learning simulation in Figures 3 and 4, respectively).

The critical difference can be seen in the error-rate just after this eight-session phase of 8-sec. learning. In the sequential learning simulation (Fig. 3), when the network is switched back to training on the 40-sec. duration, the error rate shoots up to 0.7 and remains above 0.4 for a number of sessions. On the other hand, when the 8- and 40-sec. durations were learned concurrently (Fig. 4), there is no such jump in error-rate. The error-rate for 40-sec. learning after the eight-session 8-sec. training is essentially what it was at the end of the concurrent learning phase. In short, the network now has no trouble reviving its prior memory of the 40-sec. duration.

This is very similar to what was observed in the peak-time experiments with the rat. Concurrent learning

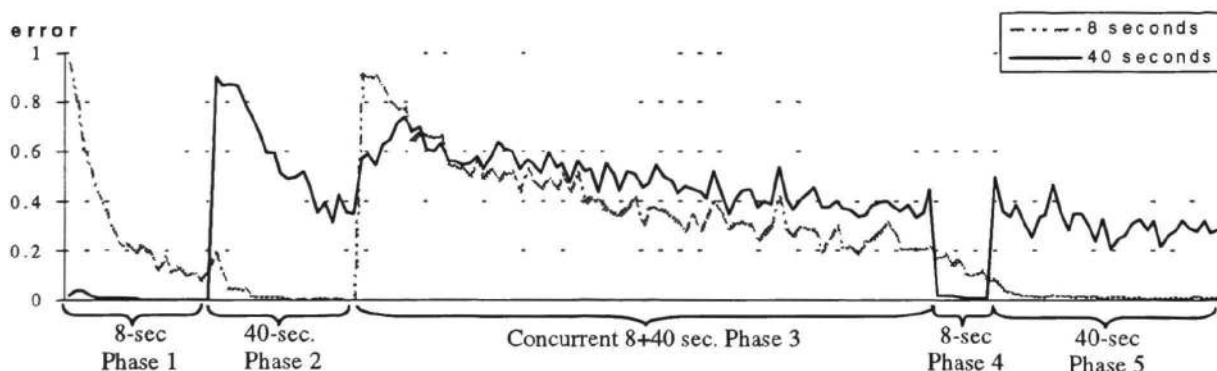


Figure 4. Concurrent learning of 8- and 40-sec. durations in Phase 3.

allows the rat to rapidly “revive” its memory trace of the 40-sec. duration; sequential learning does not. In the latter case, both in the case of the simulation and the rat, the memory trace seems to have been catastrophically erased by new learning.

Internal representations developed by sequential versus concurrent learning

Given the results of these simulations, we wondered whether the differences in forgetting observed in sequential and concurrent learning were related to the network’s internal representations of the time information it had learned. This suggests the corresponding question for the rat — namely, are the differences observed for the rat’s sequential and concurrent time-duration learning based on its development of different internal representations of these durations depending on how it learned them?

To study the internal representations of 8- and 40-sec. durations in our network, we presented “pure” (i.e., without noise) 8-sec. and 40-sec. time patterns to the network just before the eight-session 8-sec. only test phase began. This allowed us to record the hidden-unit activation patterns corresponding to 8-sec. and the 40-sec. inputs for the sequential-learning simulation and for the concurrent-learning simulation. We then calculated the hidden-unit representation overlap of these two patterns for both simulations. We then compared the differences in overlap of the hidden-unit encodings of 8-sec. and 40-sec. durations for the two different learning scenarios.

Our prediction was that, since there was considerably less forgetting of the 40-sec. time-duration in the concurrent-learning simulation, that the representations developed during Phase 3 of the concurrent-learning simulation would overlap less and be more sparse than the corresponding representations developed during sequential learning. This prediction was confirmed by the data (Fig. 5).

Hidden-layer representational overlap and sparseness

We calculated representational overlap by means of an inner-product measure of the hidden-unit vectors corresponding to 8-sec. and 40-sec. input. These values were averaged over 60 runs of the program. As predicted, in the case of case of concurrent learning, this value (0.61) was considerably lower than when the representations had developed during sequential learning (1.49). In other words, the internal representational overlap of the two time-duration patterns was *2.4 times higher for sequential learning*, accounting, at least in part, for the greater interference observed in sequential learning.

Representational sparseness can be measured by considering the total amount of activation used by a representation and by its dispersion. In general, the more sparse the representation, the less activation it will use and the lower its dispersion. For sequential learning, the average 8-sec. hidden-unit representation used 5.5 units of activation and the 40-sec. representation, 2.4 units of activation. For concurrent learning, these figures dropped to 2.2 and 2.0

units of activation, respectively. A simple measure of dispersion (standard deviation, SD) indicated how dispersed the representations were over the 18 hidden-units. The spread of the 8- and 40-sec. representations in the case of sequential learning (0.41 and 0.12 SD) is considerably greater than the spread of the same representations in the case of concurrent learning (0.27 and 0.10 SD). Figure 5 shows graphically the degree to which the sparseness of these representations differs (the activation levels of the representations have been ordered in order to facilitate comparisons between them).

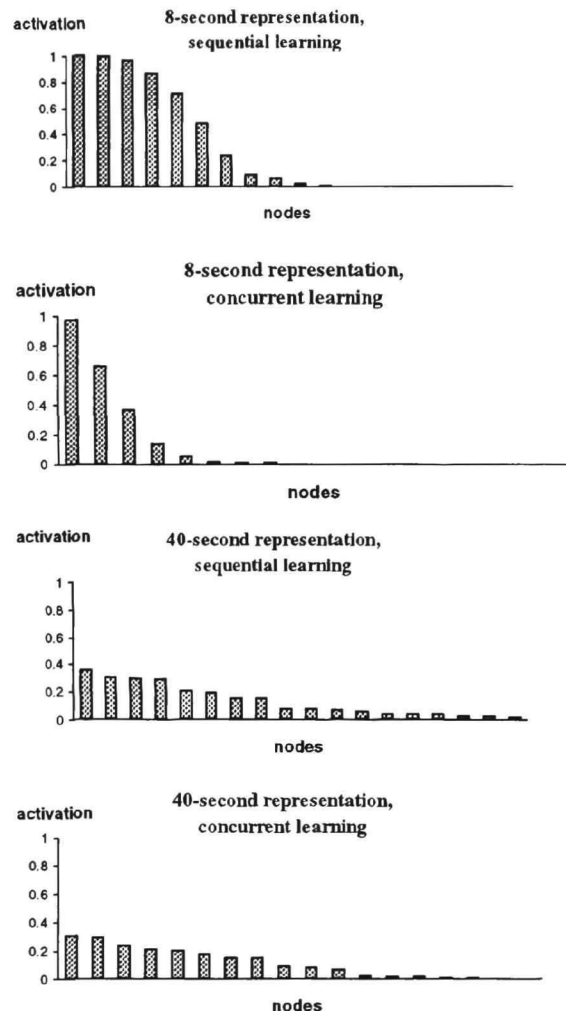


Figure 5. Differences in the sparseness of coding at the hidden layer of the representations for 8-sec. and 40-sec. durations in the case of sequential learning and concurrent learning.

In other words, the network’s internal representations of 8- and 40-sec. durations when learned concurrently, were significantly sparser and less overlapping than when they were learned sequentially. It is well established that the amount of internal representational overlap is contributes significantly to the amount of catastrophic interference produced. It is therefore reasonable to conclude that forgetting is far less catastrophic for concurrent learning than for sequential learning because of, at least in part, the

smaller amount of interference of the internal representations of the two time-durations in concurrent learning compared to sequential learning.

Conclusion

Two suggestions emerge from this research. The first is that the rat may not store time-duration information in the same way it stores other types of information that are less susceptible to catastrophic interference. In other words, the complementary hippocampal-neocortical system proposed by McClelland, McNaughton, and O'Reilly (1995) to avoid catastrophic interference may not be used by the rat for long-term storage of time information. The fact that a single connectionist network seems to produce effects quite similar to those actually observed in rats for both sequential and concurrent time-learning would argue for the possibility of a unitary time storage area in the rat, rather than a dual hippocampal-neocortical mechanism.

Secondly, we believe these results support for the claim that in the rat there may be a distinctly different internal coding of time durations when they are learned concurrently, as opposed to when they are learned sequentially. In our connectionist simulation, we obtain forgetting results similar to those observed in the rat. These differences in the network correspond to significant differences in the internal coding of the two time-durations depending on how the network learned them. The internal representations in the case of concurrent learning are more sparse and overlap less than in the case of sequential learning. We suggest that this might imply a similar type of coding in the rat.

Acknowledgments

This work was supported in part by a grant IUAP 4-19 from the Belgian National Science Foundation.

References

Barnes, J. and Underwood, B. (1959). "Fate" of first-learned associations in transfer theory. *Journal of Experimental Psychology*, 58, 97-105

Catania, A. (1970). Reinforcement schedules and psychophysical judgements, in W. Schoenfeld (ed.) *Theory of Reinforcement Schedules*, 1-42, NY: Appleton-Century-Crofts.

Church, R. & Broadbent, H. (1990). Alternative representations of time, number and rate. *Cognition*, 37, 55-81.

Ebbinghaus, H. (1885) Über das Gedächtnis: Untersuchungen zur Experimentellen Psychologie ("On memory") (H. A. Ruger and C. E. Bussenius, trans) New York: Dover, 1964

Ferrara, A. (1999). Unpublished doctoral dissertation. Psychology Department, University of Liège, Belgium.

French, R. M. (1991) Using Semi-distributed Representations to Overcome Catastrophic Forgetting in Connectionist Networks. *Proceedings of the Thirteenth Annual Cognitive Science Society Conference*, 173-178, Hillsdale, NJ: LEA

French, R. M. (1994) Dynamically constraining connectionist networks to produce distributed, orthogonal representations to reduce catastrophic interference. *Proceedings of the*

Sixteenth Annual Conference of the Cognitive Science Society, 335-340, NJ: LEA

French, R. M. (1999). Catastrophic interference in connectionist networks: Causes, consequences and solutions. *Trends in Cognitive Sciences*. (to appear).

Gibbon, J., Church, R. and Meck, W. (1984). Scalar timing in memory. In J. Gibbon and L. Allan (eds.), *Annals of the New York Academy of Sciences: Vol. 423. Timing and time perception*, 52-77, New York: New York Academy of Sciences

Hetherington, P. and Seidenberg, M., (1989), Is there "catastrophic interference" in connectionist networks?, In *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, 26-33, Hillsdale, NJ: LEA

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

Lejeune, H., Ferrara, A., Simons, F., and Wearden, J. H. (1997). Adjusting to changes in the time of reinforcement: Peak interval transitions in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 211-231.

Lewandowsky S. (1991) Gradual unlearning and catastrophic interference: a comparison of distributed architectures, in W. Hockley and S. Lewandowsky (eds.) *Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock*, 445-476, NJ: LEA

Lewandowsky, S. and Shu-Chen Li (1993) Catastrophic interference in neural networks: causes, solutions, and data, in F.N. Dempster and C. Brainerd (eds.) *New Perspectives on Interference and Inhibition in Cognition*. New York, NY: Academic Press.

McClelland, J. and Goddard, N. (1996). Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus*, 6, 654-665.

McClelland, J., McNaughton, B. and 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

McCloskey, M. and Cohen, N. (1989) Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (ed.) *The Psychology of Learning and Motivation: Vol. 24*, 109-164, NY: Academic Press

Meck, W., Komeily-Zadeh, F. and Church, R. (1984). Two-step acquisition : Modification of an internal clock's criterion. *Journal of Experimental Psychology : Animal Behavior Processes*, 10, 297-306.

Murre, J. (1992b) *Learning and Categorization in Modular Neural Networks*. Hillsdale, NJ: LEA

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

Roberts, S. (1981). Isolation of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*, 7, 242-268