pattern of abnormalities in STC mirror neurons and ACC spindle neurons. Theoretical translation of these deficits into specific schizophrenic symptoms would then require revision if neural abnormalities tied to the deficits are not consistently found in people who exhibit the symptoms.

In conclusion, recent research has built a persuasive case for linking cognitive deficits to schizophrenia, but the evidence is, in essence, circumstantial. Fine-grained analysis of specific neural substrates will provide us with even more persuasive 'eyewitness' testimony.

References

1 Sharma, T. and Harvey, P., eds (2000) *Cognition in Schizophrenia*, Oxford University Press

2 Kuperberg, G. and Heckers, S. (2000) Schizophrenia and cognitive function. *Curr. Opin. Neurobiol.* 10, 205–210

3 Harvey, P. and Keefe, R. (2001) Studies of cognitive change in patients with schizophrenia following novel antipsychotic treatment. *Am. J. Psychiatry* 158, 176–184

4 Andreason, N. *et al.* (1999) Defining the phenotype of schizophrenia: cognitive dysmetria and its neural mechanisms. *Biol. Psychiatry* 46, 908–920

5 Braver, T. *et al.* (1999) Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biol. Psychiatry* 46, 312–328

6 Allman, J. *et al.* (2001) The anterior cingulate cortex: the evolution of an interface between cognition and emotion. *Ann. New York Acad. Sci.* 935, 107–117

7 Duncan, J. and Owen, A. (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483

8 Lahti, A. *et al.* (1995) Ketamine activates psychosis and alters limbic blood flow in schizophrenia. *NeuroReport* 6, 869–872

9 Mega, M. and Cummings, J. (1997) The cingulate and cingulate syndromes. In *Contemporary Behavioral Neurology* (Trimble, M. and Cummings, J., eds), pp. 189–214, Butterworth–Heinemann

10 Ochsner, K. *et al.* (2001) Deficits in visual cognition and attention following bilateral anterior cingulotomy. *Neuropsychologia* 39, 219–230

11 Nordahl, T. *et al.* (2001) Anterior cingulate metabolism correlates with Stroop errors in paranoid schizophrenia patients. *Neuropsychopharmacology* 25, 139–148

12 Frith, U. and Frith, C. (2001) The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* 10, 151–155

13 Pickup, G. and Frith, C. (2001) Theory of mind impairments in schizophrenia: symptomatology, severity and specificity. *Psychol. Med.* 31, 207–220

14 Happe, F. *et al.* (2001) Acquired mind-blindness following frontal lobe surgery? *Neuropsychologia* 39, 83–90

15 Riscalla, L. (1980) Blindness and schizophrenia. *Med. Hypotheses* 6, 1327–1328

16 Nimchinsky, E. *et al.* (1999) A neuronal morphologic type unique to humans and great apes. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5268–5273

17 Harrison, P. (1999) The neuropathology of schizophrenia: a critical review of the data and their interpretation. *Brain* 122, 593–624

Glenn S. Sanders
Associate Professor, Dept of Psychology, University at Albany, State University of New York, U.S.A.

***Gordon G. Gallup, Jr**
Professor, Dept of Psychology, University at Albany, State University of New York, U.S.A.
*e-mail: gallup@csc.albany.edu

Helmut Heinsen
Professor, Morphological Brain Research Unit, University of Wuerzburg, Wuerzburg, Germany.

Patrick R. Hof
Associate Professor, Fishberg Research Center for Neurobiology and Kastor Neurobiology of Aging Laboratories, Mount Sinai School of Medicine, New York, U.S.A.

Christoph Schmitz
Assistant Professor, Dept of Psychiatry and Neuropsychology, University of Maastricht, The Netherlands; and Dept of Anatomy and Cell Biology, RWTH University of Aachen, Aachen, Germany.

Attempting to model dissociations of memory

Paul J. Reber

Kinder and Shanks report simulations aimed at describing a single-system model of the dissociation between declarative and non-declarative memory. This model attempts to capture both Artificial Grammar Learning (AGL) and recognition memory with a single underlying representation. However, the model fails to reflect an essential feature of recognition memory – that it occurs after a single exposure – and the simulations may instead describe a potentially interesting property of over-training non-declarative memory.

Kinder and Shanks recently reported an interesting series of simulations using the Simple Recurrent Network (SRN) connectionist framework as a model of memory phenomena during the Artificial Grammar Learning (AGL) task [1]. However, in order for computational modeling to advance a specific theoretical perspective, it must accurately reflect the essential features of the behavior modeled. The proposed model does not capture essential properties of both recognition and AGL and thus does not provide an effective alternative to the multiple-memory-system view.

In the AGL task, participants are shown a series of letter strings that conform to an underlying set of rules and incidentally acquire the ability to make judgments about whether subsequent novel strings conform to the same set of rules. Studies of amnesic patients performing the AGL task and similar recognition memory tests have reported a dissociation such that the patients exhibit intact AGL although they are impaired at recognition memory for the same stimuli [2–4]. This dissociation suggests that AGL is supported by a non-declarative (implicit) memory system that depends on structures outside the medial temporal lobe (the brain areas damaged in amnesia). The new finding of Kinder and Shanks is that the SRN, with sufficient training, is able to produce recognition-like behavior, in addition to capturing the phenomena of AGL. They report that this recognition-like learning is reduced when lower learning rates are used in the model and suggest that the model might thus capture the key aspect of the declarative/non-declarative memory dissociation using a single memory system. Two additional simulations are also reported that attempt to extend the logic of these AGL simulation studies to an examination of recognition and priming phenomena. For the simulation studies, the critical question is whether the Kinder and Shanks model truly captures the recognition part of the dissociation. If the sensitivity of the

1364-6613/02/$ – see front matter © 2002 Elsevier Science Ltd. All rights reserved. PII: S1364-6613(02)01893-4

model to previously seen stimuli is not analogous to the cognitive process of recognition, the results of their studies might instead be identifying potential features of non-declarative memory rather than demonstrating a behavioral dissociation between the two types of memory.

The SRN was originally shown by Cleeremans and McClelland to be able to predict the structure of event sequences for stimuli in AGL and related tasks [5]. It was proposed to be a model of implicit learning in AGL and when the neurological dissociations were reported by Knowlton and Squire, it was hypothesized that the SRN reflected the learning mode of the unknown area(s) of the brain that support non-declarative memory in this task. The SRN learns at a fairly gradual rate, extracting the sequential contingencies across stimulus elements through exposure to many examples. This gradual learning of element associations is what supports generalization of the model to novel stimulus elements, which is necessary to exhibit successful discrimination of novel grammatical and nongrammatical items in the AGL. The model was not originally proposed to capture the phenomenon of recognition of previously seen stimuli, as Kinder and Shanks have now proposed.

Challenges in modeling recognition with connectionism

The new element introduced in the simulations of Kinder and Shanks is the claim that the SRN can capture the phenomenon of recognition. However, human recognition memory is fundamentally a rapid acquisition of novel information in that the new memory can be acquired after a single exposure. Simple connectionist models of recognition have typically not been particularly successful owing to the phenomenon of 'catastrophic interference', in which rapid acquisition of a new memory disrupts the representations of a large subset of previously acquired memories. The theoretical review of McClelland, McNaughton and O'Reilly made the case that the information-processing demands of generalizing and reproduction systems were very different and suggested the need for a separate fast, hippocampal learning system and a slow neocortical learning system [6]. The two systems those authors described map fairly directly onto the multiple-memory-systems view hypothesized on the basis of neuropsychological dissociations seen in amnesic patients and in the AGL task. The Kinder and Shanks model attempts to argue against this idea by describing a single model to do both tasks.

The flaw in the simulation described by Kinder and Shanks is seen in their note that the training required by the model to produce this phenomenon takes '100 learning epochs'. This indicates that in order for the model to demonstrate the behavioral profile reported, it was necessary to expose the model repeatedly to the stimulus set 100 times (with stimulus items interleaved, because non-interleaved training leads to catastrophic interference). Thus, Kinder and Shanks attempt to take advantage of interleaved training and gradual learning to extract AGL knowledge and simultaneously claim to capture recognition memory. Although multiple epochs of training is a common technique for demonstrating the possibility of certain distributed representations, acquisition speed and number of exposures to the training stimuli is a fundamental characteristic of the two memory systems the model attempts to capture. Declarative memory can be as fast as a single trial and non-declarative memory is typically slow and accrues over many study trials (for category and skill learning tasks, which appear to be supported by different mechanisms than priming or classical conditioning).

The SRN is an example of a slow distributed-learning system that is capable of generalization to novel stimuli and more naturally captures the slower type of non-declarative memory that supports AGL. In the report of Cleeremans and McClelland [5], the SRN is shown to be able to acquire AGL in 'real time', that is, the amount of training the model requires to match human performance is the same as the amount of training the participants in the experiment received. Learning curves are not shown, but based on the reported results of the effect of reducing the learning rate, it is highly likely that the SRN model will exhibit successful AGL performance much sooner than exhibiting the ability to identify previously seen strings. By contrast, we would expect that human recognition memory would be intact after a single exposure to a training item, whereas AGL would require exposure to a number of AGL training stimuli.

The model proposed by Kinder and Shanks simply does not provide an account of human recognition memory. The SRN was not designed to match human recognition performance and does not capture the basic phenomenon of declarative memory: one-trial learning. Under a simulation of 'damage' (e.g. reduced learning rate), it is therefore unsurprising that the model is much more robust in the ability to capture the phenomenon of AGL compared with the recognition-like behavior for which the model was not designed.

Exact reproduction with the SRN

The SRN model is designed to gradually extract the contingencies between elements of sequences rather than be used to identify exact sequences previously seen. Thus, it is not surprising that it takes much more extra training to produce 'recognition' and that this aspect of the model is much more sensitive to reductions in learning rate. Rather than suggesting that this feature models human explicit recognition memory, this type of learning in the model might be a demonstration of a hypothesized phenomenon of compensation for declarative memory impairment by overtraining non-declarative memory until exact reproduction of training items is possible. In a study of remediation of amnesia by Glisky, Schacter and Tulving, patients with deficits in memory were training on a simple programming task by being given a large number of exposures to a training set until they could accomplish the task [7,8]. The patients were able to produce a replication of the training items, but notably also demonstrated 'hyperspecificity' in their ability to apply their knowledge suggesting that rather than developing a flexible, declarative representation of the task, they had acquired a different non-declarative representation that was relatively inflexible. The idea that non-declarative memory is more closely tied to the original learning situation and is less flexible has been proposed as a major functional difference between declarative and non-declarative memory [9,10]. This idea has been explored in additional studies of non-declarative memory and the flexible application of knowledge [11–13]. The recognition-like behavior of the Kinder and Shanks model seems more likely to be capturing the phenomenon of 'hyperspecific' non-declarative memory rather than being a model of human recognition memory. Although the Kinder and Shanks model can identify previously seen stimuli, it is bound by the sequential presentation of the stimuli and certainly couldn't produce pattern-completion-like behavior (filling in missing details) or retrieval of related instances which are basic components of flexible, explicit human recognition memory.

Behavioral evidence of interactions

Kinder and Shanks additionally report two AGL experiments that examine the effect of instructions cuing healthy participants to rely on implicit or explicit learning in order to support their single-memory-system hypothesis. However, their experimental evidence is based on finding a null result in support of their hypothesis (i.e. that the instructions didn't affect performance), and thus is not an informative design. Even with a better design, demonstrating that there is some interaction between non-declarative and declarative memory systems isn't enough to prove that there is a single system. Although there have been reports of striking dissociations between recognition and priming [14,15], there are other tasks in which declarative and non-declarative memory interact. For example, the habit-learning system of the basal ganglia appears actively to inhibit the declarative memory system in the medial temporal lobe [16]. In the AGL task, patients with Alzheimer's disease appear to be influenced by grammaticality in their recognition judgments although these patients otherwise display little evidence for declarative memory of study items (P.J. Reber *et al.*, unpublished data). Observing interactions among memory systems could be very informative about the organization of memory throughout the brain, but it will not indicate whether there are single or multiple memory representations.

Conclusion

The principal limitation of the Kinder and Shanks simulation is that it does not effectively model declarative memory (recognition) and therefore cannot show that their SRN model accounts for a dissociation between declarative and non-declarative memory with a single representation. Instead, their simulation is more likely to describe a phenomenon within non-declarative memory in which extensive training can enable identification of previously seen stimuli. This idea has been previously proposed [7,8] and would predict that the identification ability would be highly specific to the training items, which is generally consistent with the limited input and output representations of the SRN.

The general approach of using computational models to capture behavior can be very effective at uncovering the component cognitive processes in complex cognitive tasks. However, attempts to differentiate between theories of multiple or a single memory system using a computational model need to be applied with caution. A crucial part of the logic of this approach is that if a single system can demonstrate a dissociation, then the single-memory system is necessarily preferred over a multiple-memory system. However, the goal is to account for human memory function and so modeling approaches should aim to reflect as much evidence as possible from behavioral studies, neuropsychology and functional neuroimaging. Being too selective in the data incorporated into a simulation can easily result in a misleading example of dissociation by failing to account for important aspects of the cognitive functions involved.

The bulk of the evidence from neuropsychology and neuroimaging suggests multiple representational systems that reflect separate types of memory in the brain. However, the interactions (or lack of) between these systems and their information-processing aspects are not yet understood. Models like the one proposed by Kinder and Shanks that explore similarities in function and/or representation of the systems are likely to be necessary steps toward a comprehensive model of memory function. The successful and comprehensive memory-system model eventually found is likely to describe both surprising dissociations and interacting components (in the healthy brain) of the neural networks that support human memory.

References

1 Kinder, A. and Shanks, D.R. (2001) Amnesia and the declarative/non-declarative distinction: a recurrent network model of classification, recognition, and repetition priming. *J. Cogn. Neurosci.* 13, 648–669

2 Knowlton, B.J. *et al.* (1992) Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychol. Sci.* 3, 172–179

3 Knowlton, B.J. and Squire, L.R. (1994) The information acquired during artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 79–91

4 Knowlton, B.J. and Squire, L.R. (1996) Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 169–181

5 Cleeremans, A. and McClelland, J.L. (1991) Learning the structure of event sequences. *J. Exp. Psychol. Gen.* 120, 235–253

6 McClelland, J.L. *et al.* (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. *Psychol. Rev.* 102, 419–437

7 Glisky, E.L. *et al.* (1986) Computer learning by memory-impaired patients: acquisition and retention of complex knowledge. *Neuropsychologia* 24, 313–328

8 Glisky, E.L. *et al.* (1986) Learning and retention of computer-related vocabulary in memory-impaired patients: method of vanishing cue. *J. Clin. Exp. Neuropsychol.* 8, 292–312

9 Cohen, N.J. (1984) Preserved learning capacity in amnesia: evidence for multiple memory systems. In *Neuropsychology of Memory* (Squire, L.R. and Butters, N., eds), pp. 83–103, Guilford Press

10 Squire, L.R. (1994) Declarative and non-declarative memory: multiple brain systems supporting learning and memory. In *Memory Systems* (Schacter, D. and Tulving, E., eds), pp. 203–232, MIT Press

11 Eichenbaum, H. *et al.* (1989) Further studies of hippocampal representation during odor discrimination learning. *Behav. Neurosci.* 103, 1207–1216

12 Eichenbaum, H. *et al.* (1990) Hippocampal representation in place learning. *J. Neurosci.* 10, 3531–3542

13 Reber, P.J. *et al.* (1996) Dissociable properties of memory systems: differences in the flexibility of declarative and non-declarative knowledge. *Behav. Neurosci.* 110, 861–871

14 Hamann, S.B. and Squire, L.R. (1997) Intact perceptual memory in the absence of conscious memory. *Behav. Neurosci.* 111, 850–854

15 Stark, C.E.L. and Squire, L.R. (2000) Recognition memory and familiarity judgments in severe amnesia: no evidence for a contribution of repetition priming. *Behav. Neurosci.* 114, 459–467

16 Poldrack, R.A. *et al.* (2001) Interactive memory systems in the human brain. *Nature* 414, 546–550

Paul J. Reber

Dept of Psychology, Northwestern
University, 2029 Sheridan Road, Evanston,
IL 60208, USA.
e-mail: preber@northwestern.edu