Double Dissociation in Artificial Neural Networks: Implications for Neuropsychology

John A. Bullinaria & Nick Chater
Department of Psychology, University of Edinburgh
7 George Square, Edinburgh EH8 9JZ, U.K.
johnbull@uk.ac.ed nicholas@uk.ac.ed.cogsci

Abstract *

We review the logic of neuropsychological inference, focusing on double dissociation, and present the results of an investigation into the dissociations observed when small artificial neural networks trained to perform two tasks are damaged. We then consider how the dissociations discovered might scale up for more biologically and psychologically realistic networks. Finally, we examine the methodological implications of this work for the cornerstone of cognitive neuropsychology: the inference from double dissociation to modularity of function.

1. Introduction

Cognitive neuropsychology aims to inform theories of normal cognitive function by looking at how the cognitive system breaks down in patients with brain damage. The inference from patterns of breakdown to normal function is, however, notoriously difficult and such inferences depend on the theories of normal function under consideration (Gregory, 1961; Shallice, 1988; Caramazza, 1984). The methodology of cognitive neuropsychology is rooted in "box and arrow" cognitive models, in which the architecture of the cognitive system is specified in very broad terms. Patterns of breakdown are assumed to correspond to selective damage to specific boxes and arrows. Conversely, observed patterns of deficit are used to constrain how such box and arrow models should look. The augmentation of the "box and arrow" models with artificial neural network models (ANNs) of a wide range of the cognitive processes that neuropsychology has studied thus poses the question: how, if at all, should neuropsychological methodology respond to the introduction of connectionist modelling techniques? It is this issue that this paper addresses.

We begin by considering the logic of cognitive neuropsychological inference in quite abstract terms, and then concentrate on a specific methodological principle, the inference from double dissociation (DD) to modularity of function. DD has been of central importance because it promises to allow the neuropsychologist to map out the structure of the cognitive system. We review past work on the reliability of this inference for box and arrow models and in ANN models. We then present a range of simulations which show DDs between rule and sub-rule performance in small feedforward ANNs. The generality of this work is considered and we suggest that some types of damage can be extrapolated more confidently than others from lesion studies on small scale ANNs to patterns of breakdown that can be expected in the brain. Finally, we examine the methodological implications of ANN models for cognitive neuropsychology.

2. The logic of neuropsychological inference

To elucidate the nature of neuropsychological inference, we first consider the ideal conditions for such inference, and then consider what simplifying assumptions must be made in practice, when such conditions do not generally hold.

In the ideal case, predictions concerning likely cognitive deficit can be derived if the cognitive system is understood (i) in terms of the computations being performed, (ii) how those computations are implemented in the brain, and (iii) if the damage suffered is known in detail (see Caramazza, 1986; Shallice, 1988 for other discussions of the logic of neuropsychological inference). Given these prerequisites, it is possible to predict the cognitive deficits associated with each pattern of damage, compare these predictions with observed cognitive deficits, and revise conjectures about (i), (ii) and (iii) accordingly. In cognitive neuropsychology, interest focuses on the revision of (i), the computational theory of the cognitive system.

In practice, however, knowledge of (i), (ii) and (iii) is conjectural, and specified only in the broadest terms. Regarding (i), the cognitive system is often

* Research supported by the United Kingdom Joint Councils Initiative in Cognitive Science/HCI, Grant no: SPG 9029590

specified only at the level of large scale architectural organization, typically in the standard "box and arrow" notation. Recently, rather more detailed ANN models have also been considered. Regarding (ii) the neural implementation of cognitive processes is generally not explicitly considered at all, apart from some considerations of cerebral localization, largely because detailed information is not available. Regarding (iii), lesions can only be identified at a gross level, and damage is often diffuse. Since (i), (ii) and (iii) are known in such little detail, direct predictions of likely patterns of cognitive deficit cannot be derived and compared with known neuropsychological deficits. How, then, can neuropsychological data constrain cognitive theory?

A bold, but perilous, path is to make strong simplifying assumptions concerning (i)-(iii) in order to obtain predictions concerning likely patterns of damage. For "box and arrow" models, the key assumption is that brain damage selectively affects particular "boxes" and "arrows"; furthermore, it is assumed that impaired performance directly reflects the operation of this damaged system, and is not complicated by compensatory cognitive strategies. A potential problem is that even given this assumption it may not be clear what predictions can be made, unless the boxes and arrows account is specified in detail (Seidenberg, 1988). In ANN models, the crucial simplifying assumption is that brain damage can be modelled as involving the removal of, or disturbance to, particular units and/or weights. Given this assumption, it is possible to derive detailed, quantitative predictions (e.g. Patterson, Seidenberg & McClelland, 1989; Hinton & Shallice, 1989; Plaut & Shallice, 1991).

It is now clear how neuropsychological data can help decide between alternative cognitive level accounts T_1, T_2... T_n: their respective predictions P_1, P_2... P_n concerning expected patterns of damage are derived, given the necessary simplifying assumptions, and compared with neuropsychological data D. The degree to which one theory T_k is favoured over the rest depends on: (1) how well P_k matches D; (2) how well the other theories predict D.

Below, we concentrate on an aspect of patient data which has been viewed as central to cognitive neuropsychology: double dissociation.

3. The double dissociation inference

The cornerstone of cognitive neuropsychology is the inference from DD (Teuber, 1955) to modularity of function. Two tasks A and B doubly dissociate across a patient population if there are some patients who have normal or near normal performance on A, but impaired performance on B, and others with the reverse deficit. The DD inference takes this pattern of deficits to imply that A and B cannot be subserved by the same cognitive machinery. More strictly, although tasks A and B may to some extent draw on the same aspects of the cognitive system, there must be parts specific to A and others specific to B; in "box and arrow" terms, at least some box or arrow must be specific to each of A and B.

In terms of the earlier discussion, the validity of the inference from DD to a particular theory T_k of the modular organization of the cognitive system under study depends on (1) how well T_k predicts the DD; (2) how well the other theories predict a DD. The validity of (1) and (2), and hence how well DDs can distinguish between rival accounts of the functional organization of the cognitive system, depends on the class of theories T under consideration. Let us start by assuming that T includes "box and arrow" models, and then consider the case where T also includes ANNs.

3.1 Boxes and arrows.

Any "box and arrow" model in which some component is selectively used for A and another which is selectively used for B can predict a DD given the standard assumption that brain damage can cause selective damage to a particular box or arrow. Thus point (1) is straightforward.

Point (2), however, is less clear cut. Firstly, many different modular architectures can lead to the same DD. All that is required is that there is some specific component for each task. That there is such a component says nothing about its function, nor how it fits into the rest of cognitive system. For example, it is *prima facie* consistent with the DD between long and short term memory that memory consists of a very large and complex array of modules, all shared between short and long term memory, except for two, one of which has some function specific to remembering information over long periods and one which has some function specific to remembering information over short periods. Secondly, DDs between two tasks can occur even when there is no specific dedicated module for either task (Dunn & Kirsner, 1998; Shallice, 1988; see Chater & Ganis, 1991 for a very simple illustrative example).

Claims concerning what can be learnt from DDs are often put much more strongly than this. For example, Marin et al. (1976) state that: "At the very least... [observed double dissociations] ... should yield a taxonomy of functional subsystems. It may not tell us how these subsystems interact - but it should identify and describe what distinct capacities are available..." (pp 869-870). That is, they argue that DDs should specify the components of a "box and arrow" model of a cognitive system. As we have seen, such claims are not justified, even if consideration is limited to modular systems.

3.2 Neural networks.

Since the DD inference is intended to map out, or at least constrain, the architecture of the cognitive system in terms of "boxes and arrows" it might seem that ANN models are necessarily irrelevant to this aspect of neuropsychological methodology. ANN models, the argument might go, are at a level of detail below that of the box and arrow diagram which DDs are not used to uncover. This suggests that cognitive neuropsychology can proceed without concern for ANN models of cognition. The reason that this line of argument is not convincing is that it does not consider the possibility that a single ANN, without any obvious "box and arrow" structure, might be able to produce DDs. This could mislead the cognitive neuropsychologist into postulating a modular structure where none was present.

So, for example, ANN approaches have frequently aimed to model rule-governed and rule-exceptional behaviour in using a single network, where an obvious "box and arrow" model treats these as separate (see Rumelhart & McClelland, 1986; Pinker & Prince, 1988; for discussion of the past tense; Seidenberg & McClelland 1989; Coltheart et al., 1992; for discussion of reading). Hence, in terms of the above discussion ANN models can amount to new theories T concerning normal function. From the point of view of neuropsychological inference, the crucial question is what predictions P do such models make about patterns of breakdown. Specifically, can a "single route" model of rule-governed and rule-exceptional behaviour give rise to DDs? If so, the inference from DD to modularity of function is threatened; if not, the traditional inference is not challenged by ANN accounts. We discuss this question and examine a relevant case study below.

Wood (1978) and Sartori (1988) give simple demonstration simulations which show dissociation-like effects on simple pattern association tasks. Shallice (1988: 254), however, argues that these cases are not persuasive, since mere associations rather than independent tasks are dissociated and because the experiments are very small scale. Furthermore, he argues that the small scale of these experiments means that individual units and connections play an important role in the functioning of the whole system and notes that this is unlikely to be true in more realistic ANNs. He concludes that "there is as yet no suggestion that a strong DD can take place from two lesions within a properly distributed network".

In the light of these complexities it is clear that the reliability of inferences from DD to a particular functional modularity cannot be assessed purely in the abstract. We therefore consider a case study in which small ANNs are trained on a pair of tasks, systematically lesioned and examined for evidence of dissociation between the tasks.

4. Neural Network Simulations

We begin by outlining some of the problems encountered in ANN simulations in general. We then describe our models and present some typical results. Finally, we consider the important problem of scaling up to more realistic networks.

4.1 General Remarks.

ANN models are vastly oversimplified with respect to real brains, both at the level of the operation of single cells, and the patterns of connectivity between cells. The relevance of ANN simulations for neuropsychology depends on the assumption that these simplifications are not crucial with respect to the effects of damage; the effects are assumed to be the similar for any network-like system. It is not, however, currently clear even that different kinds of ANN produce similar patterns of damage. This ties in with the general problem of the parameter dependence of ANN simulations, and sensitivity to the precise weight start values. Furthermore, very different networks may be produced by different learning algorithms; one might, for example, expect that modular structures are more likely to arise from constructive algorithms (e.g. cascade correlation) than gradient descent algorithms. This issue is particularly important since no current ANN learning algorithms are biologically plausible.

A further important design question is whether the ANN is minimal, i.e. whether it has the minimum number of units and connections required to solve the problem. Minimality tends to speed up the training, improve generalization and make it easier to understand the hidden unit representations. However, minimal networks will not be fully distributed - the influence of each unit or connection will not be small. Presumably the brain has many spare hidden units, which raises the concern that imposing minimality may force the network to find solutions very unlike those found in more natural, non-minimal conditions.

There is also a dependence on the representation of training data. Many systems use complicated representations and there is much scope for 'cheating'. Often we have to encode frequency effects into the training data (e.g. word frequencies in reading models) and it is not clear how to do this effectively. We often have to present the exceptions more frequently in order for the network to learn them and we have to ask whether this should this be considered 'cheating'.

Once a particular network has been chosen and trained, the many possible types of damage must be considered. The most obvious is the removal of subsets of units and connections. Other possibilities include changing the weights and activations: adding noise, random rescaling, global rescaling, clipping weights or activation functions, and so on.

Neuropsychological patients often (but not always) show rapid improvement in performance after a lesion occurs (Geschwind, 1985). When working with minimal networks, we can easily lesion them so that they become sub-minimal. In these cases, one has the option of allowing relearning after damage. This can further confuse the results: relearning can create, destroy or even reverse the sense of dissociations. For non-minimal networks, we do not have this problem: the relearning invariably totally compensates for the damage and we get no dissociations at all. In summary then, there are a number of reasons why interpretation of ANN simulations of lesion damage is very difficult.

4.2 Simulation Results.

We trained a range of small feed-forward ANNs, with one hidden layer, on semi-regular mappings (involving a rule and a less frequent sub-rule). The networks were then lesioned in a variety of ways. The frequencies of errors on each pattern were counted, and compared with the numbers of rule and sub-rule errors expected by chance. The statistical significance of the difference was measured using chi-square tests. We found that dissociations were surprisingly common in populations of nets and that DDs could also be found within a single network. This appears to reinforce doubts regarding DD (e.g. Dunn & Kirsner, 1988; Chater & Ganis, 1991), but a more detailed investigation suggests otherwise.

The following table shows the strongest dissociations found for a typical network, with 8 inputs, 16 hidden units and 8 outputs, trained using a conjugate gradient algorithm. The training data consisted of the identity map except that when the first four input bits are 1111 or 0000 the last three bits are flipped. The full set of 256 training patterns was used, giving 224 'rules' and 32 'sub-rules'; each sub-rule pattern was presented twice per epoch (Table 1).

Form of damage	Rule errors	Sub-rule errors	*p* value
Scaling weights - globally	0.0%	96.9%	$< 10^{-6}$
Scaling weights - randomly	37.1%	100.0%	$< 10^{-6}$
Shifting weights - noise	21.4%	84.4%	$< 10^{-6}$
Removing hidden unit 1	0.9%	50.0%	$< 10^{-6}$
Removing hidden unit 2	50.0%	25.0%	< 0.008
Removing I-H link 1-3	0.4%	50.0%	$< 10^{-6}$
Removing I-H link 2-4	41.5%	0.0%	$< 10^{-5}$

Table 1. *Damaging a backprop rule/subrule net*

Similar results were obtained when the same problem was solved using a constructive algorithm, a variant of Cascade Correlation (Fahlman & Lebiere, 1988), (Table 2).

Notice that although DDs are found, they are quite weak, especially the dissociations where the rules are lost (i.e. these are more likely to occur by chance). Also although there are twice as many hidden units as in a minimal network for this problem, there are still some hidden units and connections that *on their own* have such an influence on the outputs that their removal gives rise to a dissociation. Hence, according to Shallice's criterion, noted above, these networks are not fully distributed. This has important implications for networks damaged by the removal of random subsets of units and connections. With many units and connections having very little effect on the outputs it will be quite common to find dissociations that arise due to a very small number of crucial units which have a significant effect on the output, but do not perform any identifiable function on their own. In particular, they are not performing a function revealed by the observation of a DD.

Form of damage	Rule errors	Sub-rule errors	*p* value
Scaling weights - globally	0.9%	100.0%	$< 10^{-6}$
Removing hidden unit 1	8.0%	50.0%	$< 10^{-6}$
Removing hidden unit 2	50.0%	18.8%	$< 10^{-3}$
Removing I-H link 1-3	2.2%	53.1%	$< 10^{-6}$
Removing I-H link 2-4	39.7%	0.0%	$< 10^{-5}$

Table 2. *Rule/subrule cascade correlation net*

An unexpected feature of the results is that errors on a pattern do not necessarily occur where expected. For example, errors on sub-rule patterns sometimes occur on parts of the input string where the mapping is completely regular.

There was also evidence that the pattern of dissociations is very task-dependent. With the above training data we can find dissociations in the number of bit errors with completely random weights (where we expect 50% errors for both rules and sub-rules) much more frequently than we would expect by chance (calculated by chi-squared), (Table 3). The pattern of effects remains much the same if very much sparser training data is used - just 1.5% of possible patterns.

		Full data set	1.5% of data set
Instances occurring with expected probability	<= 1	10000	10000
	$<10^{-1}$	2491	1897
	$<10^{-2}$	806	432
	$<10^{-3}$	293	95
	$< 10^{-4}$	85	21

Table 3 *Expected and actual number of dissociations between rule/subrule performance*

On the other hand, if we use a training set that has rules and sub-rules specified by a different procedure (namely a parity rule) we can find far *fewer* dissociations than expected by chance (Table 4).

		% training data used				
		all	25%	25%	6.5%	6.5%
Instances occurring with expected probability	<= 1	10^4	10^4	10^4	10^4	10^4
	$<10^{-1}$	12	1262	237	1717	458
	$<10^{-2}$	0	138	2	307	18
	$<10^{-3}$	0	11	0	51	3
	$< 10^{-4}$	0	0	0	1	1

Table 4 *Lesioning a parity network*

Thus the number of dissociations appears to depend crucially on the task used.

4.3 Scaling up.

We have found DDs in ANNs but it is not clear whether DDs can occur in larger and more distributed networks. Unfortunately scaling up presents a number of difficulties, and exacerbates many of the problems mentioned in section 4.1 For example, is the number of hidden units and layers sufficient to allow the networks to solve the problem in a natural modular manner, or are they forced to operate in an unnatural manner? Do we have enough training patterns to prevent the network from operating by table lookup? Suppose we had succeeded in training a network to perform basic arithmetic. It would have to be quite large and large networks are very difficult to analyse in detail. It is quite likely that it would do single digit additions and multiplications by table lookup, it might have module-like components to do long additions/subtractions making use of these tables, and so on. How, then, could we decide if it had developed separate "modules" for long multiplication and long division. Using the DD methodology, we would look for forms of damage such that long division was lost but not long multiplication, and vice versa. For concreteness, suppose that the two modules each consisted of 100 units and the rest of the system was another 200 units. For very small amounts of artificial lesion damage, it is possible that one system would be selectively damaged and the other preserved; but for larger amounts of damage, this would become almost infinitesimally unlikely; and a combinatorialy explosive number of possible lesions would have to be performed to uncover such dissociations. So, even if there is modular structure present in ANNs, large scale models with large scale damage, are unlikely to give rise to dissociations. Furthermore, the ANNs would be almost as difficult to analyse as brains.

Notice, though, that if biological learning algorithms tend to organize neurons with common function into local brain regions, or such localisation is enforced by innate constraints, then the chances of lesion damage affecting one task selectively, resulting in a dissociation, increases significantly. This is one reason why current ANNs may provide unreliable models of neuropsychological breakdown.

ANNs may model more global kinds of damage more successfully. For example, neurotransmitter imbalances can be crudely modelled by globally rescaling weights, which can easily be tested on ANNs however large. We have found no evidence that this kind of damage can give rise to DDs; in the tasks reported above, the subrules/exceptions are generally lost and the rules spared.

5. Implications for neuropsychology

In this section, we consider the implications of these results for cognitive neuropsychology.

5.1 Do double dissociations specify modularity?

We have shown that DDs can arise in simple ANNs in a rule/sub-rule learning task. How useful such results are depends on whether we are concerned to show that: (1) ANN models are consistent with what goes on in real brains; (2) DDs are possible in fully distributed systems and consequently cannot be used to infer modularity; or (3) modular structures can arise spontaneously by learning in a fully distributed system.

If one is just interested in (1) then the question of modularity is irrelevant: the workings of our models can simply be as mysterious as those of real brains. Cases (2) and (3) are more subtle. As noted in Section 4.3, in any ANN system large enough to be considered fully distributed, it will be difficult (if not impossible) to uncover modularity without looking for DDs anyway, so even if a DD in a large scale ANN were found, it would be difficult to argue for (2) against (3). We know that a certain amount of modularity occurs in real brains, but most is clearly innate. Thus if we assume case (3), we end up simply trying to show how non-innate modular structures could arise in the brain. Moreover, the possibility of innate structures in real brains means that results from ANNs can't really tell us anything for certain. If we do find DDs, then we don't know what it implies. If we don't find DDs we don't know if it is because DDs in real brains arise solely due to innate structures that haven't been built into ANNs or because ANN learning algorithms are too dissimilar to those in real brains for the same modular structures to arise. Furthermore, as noted in Section 3.1, even if DDs could be shown to imply some modularity of function, there will still be all manner of modular and quasi-modular systems which are consistent with DD.

5.2 Methodological implications.

Despite finding DDs in ANNs, given the problems of extrapolating from small artificial simulations to real brains, one cannot really justify the suggestion that DDs are not, after all, useful data for constraining cognitive theory. Indeed, any particular DD will pose an important challenge for any non-modular ANN account; whether or not such a challenge can be met must be determined on a case by case basis. For example, single route ANN models of reading have been proposed (e.g. Seidenberg & McClelland, 1989), but cannot account for the DD between non-word and exception word reading (e.g. Coltheart et al., 1992) and this poses an important challenge for such models. Notice, however, that DD is on a par with any other aspect of neuropsychological or experimental data - it has no specially decisive importance.

The morals concerning the impact of ANN models on cognitive neuropsychology can now be drawn. First, whether a particular ANN account is consistent with a DD cannot be determined for certain a priori, but, like other experimental or neuropsychological data, must be tested by computational experiments. Second, the focus on very gross patterns of data, such as DDs, has been partly due to the fact that "box and arrow" cognitive models are not detailed enough to give more detailed predictions. The rich, quantitative predictions of fully explicit computational models, such as ANNs, give rise to a wide range of predictions (e.g. the correlation between "visual" and "semantic" reading errors, and effects of concreteness/abstractness on reading in deep dyslexia (Plaut & Shallice, 1992)), among which DDs have no special status. The connectionist neuropsychologist will be able to use more fine-grained evidence to constrain cognitive theory, thus reducing the emphasis on double dissociations.

References

Caramazza, A. (1984) The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language*, **21**: 9-20.

Caramazza, A. (1986) On drawing inferences about the structure of normal cognitive systems from analysis of patterns of impaired performance: The case of single-patient studies. *Brain and Cognition*, **5**, 41-66.

Chater, N. & Ganis, G. (1991) "Double dissociation and isolable cognitive processes." *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum, pp 668-672.

Coltheart, M., Curtis, B. & Atkins, P., (1992), Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches, submitted to *Psychological Review*.

Dunn, J.C. & Kirsner, K. (1988), Discovering functionally independent mental processes: The principle of reversed association, *Psychological Review*, **95**, 91-101.

Fahlman, S. E. & Lebiere, C., (1990), The Cascade-Correlation Learning Architecture, p524, In D.S. Touretzky (Eds) *Advances in Neural Information Processing Systems 2*, Morgan Kauffmann.

Geshwind, N. (1985), Mechanisms of change after brain lesions, *Annals of the New York Academy of Sciences*, **457**, 1-11.

Gregory, R. L. (1961) The brain as an engineering problem. In W. H. Thorpe & O. L. Zangwill (Eds) *Current problems in animal behaviour*. Cambridge: Cambridge University Press.

Hinton, G.E. & Shallice, T., (1991), Lesioning an Attractor Network: Investigations of Acquired Dyslexia, *Psychological Review*, **98**, 74-95.

Marin, O.S.M., Saffran, E.M. & Schwartz, D.F., (1976) Dissociations of language in aphasia: Implications for normal functions, *Annals of the New York Academy of Sciences*, **280**, 868-884.

Patterson, K. E., Seidenberg, M. S. & McClelland, J. L. (1989) Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed) *Parallel distributed processing: Implications for psychology and neurobiology*. Oxford: Oxford University Press.

Pinker, S. & Prince, A. (1988) On Language and Connectionism: Analysis of a Parallel Distributed Model of Language Acquisition. *Cognition*, **28**, 73-193.

Plaut, D.C. and Shallice, T., (1992), Deep Dyslexia: A Case Study of Connectionist Neuropsychology, submitted to *Journal of Cognitive Neuroscience*.

Plaut, D.C., McClelland, J.L. & Seidenberg, M.S. (1992), Reading Exception Words and Pseudowords: Are Two Routes Really Necessary?, Talk presented at the *Annual Meeting of the Psychonomic Society*, St. Louis, MO, November 1992.

Rumelhart, D. E. & McClelland, J. L. (1986) On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds) *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Vol 2, Cambridge, Mass: Bradford/MIT.

Sartori, G., (1988), From Neuropsychological data to theory and vice versa, in *Perspectives in cognitive neuropsychology* (eds. G. Denes, P. Bisiacchi, C. Semenza, E. Andrewsky), London: Erlbaum.

Seidenberg, M.S. (1988) Cognitive Neuropsychology and Language: The State of the Art. *Cognitive Neuropsychology*, **5**, 403-426.

Seidenberg, M. S. & McClelland, J. L. (1989) A distributed, developmental model of word recognition and naming. *Psychological Review*, **96**, 523-568.

Shallice, T. (1988) *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Teuber, H. L. (1955) Physiological psychology. *Annual Review of Psychology*, **9**, 267-296.

Wood, C.C., (1978), Variations on a theme of Lashley: Lesion experiments on the neural model of Anderson, Silverstein, Ritz & Jones, *Psychological Review*, **85**, 582-591.