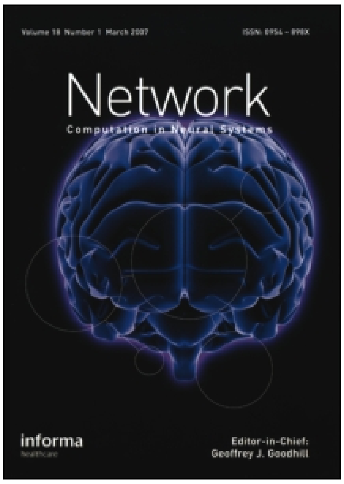


**Network: Computation in Neural Systems**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713663148

# Architecture of attractor neural networks performing cognitive fast scanning

Daniel J. Amit [a]; D. Sagi [b]; M. Usher [c]
[a] Department of Physics, London, UK [b] Department of Applied Mathematics, Weizmann Institute of Science, Israel [c] Department of Physics, Tel Aviv University, Israel

Online publication date: 01 January 1999

**To cite this Article** Amit, Daniel J., Sagi, D. and Usher, M.(1990) 'Architecture of attractor neural networks performing cognitive fast scanning', Network: Computation in Neural Systems, 1: 2, 189 — 216
**To link to this Article: DOI:** 10.1088/0954-898X/1/2/005
**URL:** http://dx.doi.org/10.1088/0954-898X/1/2/005

Network 1 (1990) 189–216. Printed in the UK

# Architecture of attractor neural networks performing cognitive fast scanning

Daniel J Amit†‖, D Sagi‡ and M Usher§

† Department of Physics, Imperial College, London, SW7 2BZ, UK
‡ Department of Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel
§ Department of Physics, Tel Aviv University, Ramat Aviv, Tel Aviv, Israel

Received 20 July 1989



**Abstract.** The salient features of the findings of high-speed scanning experiments of the Sternberg type and several attempts to account for them are reviewed, as well as the challenge implied by these findings for the general approach of *attractor* neural networks (ANNs).

We formulate a detailed, biologically flavoured, neural network, composed of three sub-networks: one preserving the test stimulus (probe), one encoding the memory set, and one encoding decision elements. Each network is a fully fed-back ANN, and serves basically to classify its inputs into classes of those inputs that lead quickly to attractors, i.e. states in which a well defined set of neurons emits bursts of spikes, and those inputs which do not. The other role of the ANN is to be able to preserve the results of the classifications for extended times. The networks communicate between themselves only when they arrive at attractors, i.e. only bursts are communicated between sub-networks. The conceptual division of the single network into sub-networks, despite the fact that the density of connections is uniform, is made possible at boundaries where the connectivity changes from full feed-back to essentially feed-forward.

The central network rehearses the memory set, on every presentation of a probe, as an ordered temporal sequence of attractors. The rehearsal is on a subliminal timescale of tens of milliseconds. If during the rehearsal the sequence goes through an attractor which is very similar to the probe, the third network records this fact in an attractor of its own, but the final reaction is conditioned on the completion of the rehearsal, which provokes the decision (logical) network into a final attractor implying either positive or negative response. Only attractors are considered meaningful events in the various component networks.

In the process of construction we confront the problem of recognition, as opposed to recall, in an ANN. This brings out an essential role that the nonlinear operation of inhibitory synapses may have in making the comparison of attractors in different networks.

The network is constructed, subject to two main constraints: the exclusive role of attractors (bursts) and the absence of grossly implausible synaptic connectivity. The construction ensures that the network exhibits the main characteristics of the experiments. The robustness of its performance is analysed and simulated. Finally, given the detailed construction of the network, of biological flavour, we speculate on the effects of possible variations in the building blocks and suggest some new experiments which its structure implies.

## 1. Experimental indications for serial search

Folk psychology suggests that we have a large-capacity memory with fast recall capabilities. Thus, as far as folk science goes, theories of memory should avoid any

‖ On leave from the Racah Institute of Physics, Hebrew University, Jerusalem, Israel.

slow processes or processes that involve a small number of memorised items. However, the introduction of reaction-time measurements into memory studies has revealed some surprising features of memory, the most interesting being the discovery of serial, time-consuming, search across memorised items, involved in retrieval processes (Sternberg 1966). In a typical Sternberg experiment subjects judge whether a test pattern (probe) is contained in a short memorised sequence of patterns (*memory set*). The subjects are asked to respond as fast as possible, using a yes or no response, but with as few errors as possible. Sternberg (1966) found that reaction time increased linearly with the number of patterns in the memory set, at a rate of about 38 ms/item, implying a serial memory search. Moreover, he found that this increase in reaction time, with the increasing number of patterns in the memory set, is the same for positive responses (probe found in memory set) and for negative responses (probe not in memory set) suggesting an exhaustive search.

This last observation seems to imply that the memory search is inefficient, since it covers the whole memory set even though on the average the probe can be found halfway through the list. The exhaustiveness of the scanning process made the scanning hypothesis less acceptable, but the robustness of the memory-set-size effect made its impact on psychologists. It seems to survive most practice schedules, unless the subjects are being trained on the same fixed memory sets over many days (Ross 1970, Kristofferson 1972b). Extended practice seems to shorten reaction time but does not affect the rate of scan, which seems to be constant (Kristofferson 1972a). Most experiments yield a rate of 30–40 ms/item when the memorised items are digits or letters, and this slope persists across different experiments and subjects, much more than the absolute reaction time. For example, schizophrenics, alcoholics and college students all have a slope between 34 and 37 ms/item while their absolute reaction time may vary across a range of 200 ms, e.g. 400–600 ms (Sternberg 1975).

The scanning rate may depend on the complexity of the items in the memory set; thus it can be as fast as 35 ms/item for digits, colours and letters and as slow as 80 ms/item for random shapes and nonsense syllables. It is interesting to note that scanning speed is inversely proportional to the mean memory span (number of items that can be stored in short-term memory) so that a total memory span may be equivalent to 245 ms of search (Cavanagh 1972). In addition, Sternberg (1975) points out that for slower scan rates a self-terminating search may be observed.

Since the original work of Sternberg (1966) many others have studied the effect of a large number of factors on performance in the item-recognition paradigm. Most of the results are described in extended reviews by Sternberg (1975) and McNicol and Stewart (1980). Many of the studies question the explanation in terms of high-speed exhaustive scanning processes. Here we relate to a small number of phenomena that seem to be common to most findings, but in no way represent the full range of effects found in scanning processes. Two of them were already mentioned, the robustness of the scanning rate and the apparent exhaustiveness. Although memory scan seems to be a phenomenon related to short-term memory, since usually the probe is presented (in a varied set procedure) only 2–3 seconds after the presentation of the memory set, similar results may be obtained in experiments where the memory set is stored in long-term memory (fixed set procedure) and the same set is used for ten minutes. In the latter case subjects can recall the sets they worked with several days later; however, the scanning process itself does not seem to be penetrable to introspection.

Some difficulties for the serial scan account come from experiments with repeated items in the memorised set. It was found that if the probe is not the repeated item then the repetition has no effect; the repeated item is counted as many times as it appears in the list in agreement with the serial exhaustive account. However, if the probe is the repeated item, the response time shortens by about 50 ms (Baddeley and Ecob 1973). Furthermore, if the scanning process were exhaustive one would expect the reaction time RT to be independent of the position of the recognised pattern in the set. Some studies, however, find serial position effects, with shorter RTs for recognitions of items placed towards the end of the memory set. But this effect seems to appear only if the probe appears less than a second after the termination of the memory set, or if the list is presented rapidly (Clifton and Birenbaum 1970).

## 2. Alternative accounts



Alternative accounts are of two classes: accounts postulating parallel processing of the memorised items and accounts postulating serial processing but self-terminating. Parallel processing accounts assume that the probe is compared in parallel to all members of the memory set. In order to explain the set-size effect these accounts usually resort to one of two features: limited capacity or stochastic variation in comparison time. One of the conclusions to be drawn from the concrete model we shall describe below is that such accounts are not necessarily exclusive.

### *2.1. Limited-capacity accounts*

It is assumed that the system has limited 'resources' which are usually left unspecified. Increasing the number of memorised items forces the system to share resources when all the items are compared in parallel. Sharing resources is then taken to imply slower processing. In its general form this account is hard to disprove by psychophysical methods since its main difference from the serial account is the temporal sequence of events during the memory scan itself. In a sense, serial search accounts may also be regarded as limited-capacity accounts. The distinction between serial and parallel here would be interesting if the accounts were able to give a theoretical justification for seriality (or parallelism) in order to make one type of search or the other essential.

An interesting attempt to put some meaning into limited-capacity accounts is the 'trace strength' account (Baddeley and Ecob 1973). It assumes the existence of a *memory trace* for each memorised item. This trace is strengthened every time an item is presented or rehearsed—the stronger the trace is the shorter the RT. The set-size effect is obtained here by limiting the total trace available; however, this account has some more interesting predictions regarding serial-position effects and repetition effects.

It should be pointed out that it is possible to construct *specific* models, with biological flavour, which have fixed 'resources' and do not slow down at all when the number of memories increases. The attractor neural network of the Hopfield kind is a case in point. See, for example, the discussion in section 3, below. When its resources are overtaxed it stops performing as an associative memory (Hopfield 1982, Amit *et al* 1987a, Amit 1989).

### *2.2. Stochastic variations in comparison time*

A stochastic accounts assumes that the comparison times of the probe to memorised patterns are independent and stochastically distributed with non-zero variance. This account can predict a set-size effect since, despite the parallel processing of the comparisons, performance is limited by the largest comparison time. This time increases with increasing number of comparisons and if the distribution is properly chosen the time may be linear in the size of the set. However, the stochasticity should manifest itself in a variance in reaction times to comparisons with sets of one item. This variance is usually too small to predict the observed set-size effect (Townsend 1971, Sternberg 1966).

### 2.3. *Self-terminating serial search*

Here it is assumed that the memory set is searched until the probe is found. For a positive memory set of size $n$, only $n/2$ items are searched on the average if the probe is in the positive set, otherwise all $n$ items must be searched. This account predicts, therefore, that positive RT as a function of $n$ will have half the slope of the negative RT. In addition, positive RTs are predicted to have a variance that increases with $n$, contrary to the exhaustive search account that may assume constant variance. (It is assumed that some noise always exists in the system, thus search time is not fixed across trial.) This is a consequence of the fact that under self-terminating search the probe maybe detected at different positions on the list while under exhaustive search it will always be detected at the end of the search. Another aspect of this same assumption is that the minimum positive RT (within the variance of a fixed $n$) will increase linearly with $n$ for exhaustive search but not for self-terminating search. This latter prediction was tested by Lively (1972) and Lively and Sanford (1972) and the data were found to be consistent with the exhaustive search account.

Another interesting variant of self-terminating search is one suggested by Theios *et al* (1973). They suggested that search is not restricted to items in the memory set but rather to all recently presented items, including those from previous trials and thus including all items belonging to the experimental set (e.g. ten items in experiments where digits are used). Items within this list are ordered by recency; thus the memory set will occupy the first $n$ positions and will be searched first. Search is terminated when the probe is found. As a result, when the probe belong to the memory set it will be found within the first $n$ steps ($n/2$ on the average), but if the probe does not belong to the memory set it will be found within the next $M - n$ steps (note that $n + (M - n)/2 = M/2 + n/2$ on the average, where $M$ is the total number of items used in the experiment). One therefore finds equal slopes for positive and negative RTs, a result that is usually attributed to exhaustive search. However, RT variance would behave the same as in self-terminating search.

### 2.4. *Hybrid accounts*

It is possible to combine features of the above accounts to obtain better results. Atkinson and Juola (1974) suggested that the probe is first compared with the items in memory in parallel; the familiarity-strength based result is then evaluated as positive, negative or ambiguous. The positive/negative results trigger an immediate yes/no response while the ambiguous result starts a serial search. Their account incorporates repetition effects as well the linear increase in RT.

Anderson and Bower (1974) suggested another hybrid account where the probe is compared to memory items by several serial searches running in parallel.

## 3. Sternberg's experiments: a challenge for ANNs

The challenge which the Sternberg experiments present to the psychologist stems from the apparent exhaustiveness of the search, decorated by some exotic phenomena of recency and repetition. For practitioners of attractor neural networks (ANNs), the challenge is significantly more fundamental. In order to bring the magnitude of the challenge into proper relief we shall recapitulate briefly the basic tenets of the ANN paradigm, within the research program initiated by Hopfield (1982, 1984). The reader who feels the need for a more exhaustive exposition is referred to Amit (1989).

An ANN is an assembly of neurons connected by synapses. The resulting network follows a dynamical course in which each neuron receives inputs from many other neurons in the assembly, based on the state of activity of the other neurons, weighted by the connecting synaptic efficacies, and then, using a nonlinear threshold function, chooses its own new state of activity. When neurons change their states of activity in a rapid asynchronous way the state of the network is followed in time by observing the list of values of instantaneous activities of all neurons. This is referred to as a network state. As neurons change their individual activity states, the network wanders from one network state to another. What determines the trajectories of the network in the space of its states is, essentially, the matrix of synaptic connections. They are supposedly also the repository of learning, of memory, or of prior organisation.

The significant 'cognitive' event in the ANN is its convergence into an *attractor*. An attractor is recognised by the fact that consecutive network states repeat. In other words, at consecutive time slots the same neurons emit spikes and the same ones remain effectively quiescent. This will take place if the pattern of activity of the network produces inputs into all neurons which give rise to the same activity of each one of them. A network state that is an attractor is interpreted as an act of recognition, or recall from memory. Which network states are attractors is again determined by the synaptic connections, which control every aspect of the dynamics. The fact that the attractors are a property of the organisation of the network, which in turn is formed by a learning process, lends plausibility to the identification of the arrival of the ANN at an attractor as a retrieval from memory.

The process of recall, and traditionally also of recognition, has been the following: some state of activity is imposed on the neurons of the ANN, or on some of them, as an expression of an external stimulus. As an example (only) consider a two-dimensional arrangement of the neurons, parallel to a retina; then a stimulus may be an activation of each of the neurons in the network in correspondence with the light intensities falling on the retina (see e.g. Buhmann and Schulten 1987). Then the network is allowed to develop under its own dynamics and its drift towards an attractor is a process of *error-correction*, *content-addressing* or the recall of the prototype network state, which is the attractor, by the stimulus.

One of the central achievements of recent modelling of neural network memories has been the discovery that a single network can have a diverse set of attractors. Which particular network states become attractors can be directly controlled by the specification of the synaptic efficacies. Once those have been specified, by learning or by fiat, one may consider the operation of the network as an associative memory for the embedded patterns which are retrieved, or recalled, by external stimuli as explained above.

The main attraction of the approach is that the dynamical process by which the network changes its states is akin to sliding down slopes in a landscape, with the attractor memories at the bottoms of valleys in the landscape (local minima). Thus, the network, whose neurons act in parallel but in complete asynchrony, can arrive very rapidly from a stimulus to the attractor it recalls. It is presumed that this asynchronous parallelism accounts for the rapid way in which biological brains recognise patterns, despite the underlying slow operation of the individual neuronal elements.

In fact, if the stimulus is not very far removed from the memorised prototype this time is:

- independent of the number of memorised patterns;
- composed of a couple of neural cycle-times required to reach the bottom of the valley in addition to the time required by a read-out mechanism to ascertain that the network has been in an attractor.

When the stimulus is quite similar to the memorised pattern the first duration shrinks essentially to zero, again independently of the number of memorised patterns. The length of the second period varies from network to network. It should extend over several cycles in which neurons can actually re-examine their state, i.e. to emit another action potential. It must therefore be a multiple of the shortest average period between spikes for the relevant neural network. This cycle may vary between 6–7 ms for certain parts of secondary visual cortex (Anderson and Mountcastle 1983) to 20–30 ms in higher associative areas. Correspondingly, the second duration may vary from 30–40 ms to as much as 100-150 ms.

What is especially pertinent for our discussion here is that the time is totally independent of the number of memorised patterns. If, in fact, recognition is the drift provoked by a stimulus into a nearby attractor, then neural networks would predict that the reaction time for recognition in the Sternberg experiments should be totally independent of the number of patterns in the memorised set. The same would be true, in this context, for non-recognition. Non-recognition can either be described into a universal attractor which attracts stimuli far from any of the learned patterns (Shinomoto 1987). In that case it acts just like another stored pattern and the previous considerations apply to it. Alternatively, non-recognition may be expressed by the fact that for a duration of the length of a cognitive event the consecutive states of the network are uncorrelated (Parisi 1986), which again is independent of the number of memories.

ANNs are consequently faced with a challenge that appears more basic than exhaustive compared with self-terminating searches. If the Sternberg experiments probe a standard elementary act of recognition then ANNs are the wrong kind of model.



## 4. Recognition in ANNs and a lingering problem

The original account of recognition in the attractor networks, as described above, should not seem quite satisfactory upon closer scrutiny. A network may arrive at an attractor for a variety of reasons which one would not want, intuitively, to label as recognition, as in the following examples.

- A stimulus is identified as unfamiliar, but provokes, by association, a well known cognitively identified image. It would be natural to associate the cognitive identification of the image as the arrival at an attractor. Yet, it is an instance of non-recognition. In other words, an attractor may have a basin of attraction which is *too wide*, and may include stimuli which one would not wish to identify with the memory.

- A stimulus gives rise to the recall of a tune. Various consecutive elements in the tune may be plausibly associated with presence in quasi-attractors. Yet, they could not be described as recognitions of any external stimuli.

Recognition must, therefore, be accounted for by a fact more complex than the mere connection of stimulus to attractor. A simple additional ingredient could be:

- Comparison of the network state of the attractor to the stimulus, following confirmation that an attractor has been reached. Recognition will be restricted to the case in which the similarity between the attractor and the stimulus is above a certain threshold. The decision as to whether an attractor represents recognition or is instead an *association*, must also be affected by attractors. See, e.g., sections 7 and 11.3.



Often the stimulus is present long enough, relative to the time of arrival at an attractor and the cognitive identification of this fact, for the comparison to be made. Otherwise, one would require an auxiliary network to preserve the stimulus for identification purposes. All these issues will naturally surface in the discussion of the substrate of the ANNs. The mechanics of comparison will be discussed in subsections 8.1 and 11.3 and the auxiliary network in subsection 11.1.

While we argue that cognitive recognition requires a more elaborate description than naively given it by the ANN community, a complication which makes certain requirements on the underlying elements (see e.g. subsections 8.1 and 11.3), the tension with the Sternberg experiments remains undiminished. Nothing in the superposed comparison introduces a dependence on the number of memorised items. Thus, if reconciliation is to be found it should be elsewhere.

## 5. Conditioned functioning and basic modelling postulate

The special features of the linear dependence of the reaction time on the size of the 'positive set' are clearly not universal and are significantly conditioned by the experimental set-up. Recognition time of items in large memory stores is evidently not linear (Standing 1973). Moreover, one would guess that recognition of items in small sets, where the fact that a set exists is not consciously prominent, will again not be linear.

Our *basic premise* is that a subject prepared for the Sternberg experiment is conditioned to rehearse the elements of the positive set in their temporal order. When a probe-stimulus is presented the subject goes over the temporal sequence and only upon its completion produces a reaction, which is an account of whether during the rehearsal a recognition has taken place. In other words, we propose that the temporal sequence composed of the positive set serves as a clock timing the reactions, while recognition takes place independently.

We point to two observations in order to motivate such a postulate. First, in early accounts of the experiment (Sternberg 1966), it is indicated that 'the trial ended with his attempt to recall the series *in order* .... Recall was imperfect on 1.4% of the trials' (emphasis added). The second observation is that the effect is reproducible only for positive sets of six items or less. This reminds us of the '$7 \pm 2$ rule' characterising phenomena of short-term memory, often associated with the recall of ordered temporal sequences (Bloom *et al* 1985).

This we do not consider as compelling arguments for our scheme but merely as circumstantial evidence which guides us to the construction of a very specific model which can then be judged on its own terms. Various possible tests will be discussed in section 10. In order to prepare for this construction we first provide a brief account of the processing of temporal sequences by neural networks. More detailed versions can be found in Amit (1988, 1989).

## 6. Robust cognitive processing of temporal sequences by ANN

The basic network organisation introduced by Hopfield (1982) was intended to endow the network with attractor dynamics, as was explained in section 3. It was based on a set of synaptic efficacies which ensured that:

(1) Each neuron received information on the immediately preceding state of the network, namely of the activity states of the other neurons in the network.
(2) The synaptic efficacies modulated the information about the activity states of the other neurons in such a way that a *variety* of prescribed patterns, expressed as specific activity distributions of neural activities in the network, would repeat themselves under the neural dynamical process.
(3) Moreover, these specific neural activity distributions, the stored patterns or the memories, would attract a wide collection of neighbouring network states.

We have identified as 'cognitive events' the presence of the network in an attractor for a long enough period, so that the wider system can ascertain the special occasion. The arrival at the attractors as well as the presence in them was shown to be very robust to a whole variety of noise, structural as well as functional (Amit *et al* 1985, 1987a, Sompolinsky 1986, 1987, Amit 1989). As the networks were originally conceived they had no provisions for the dynamics beyond the arrival at the attractor, unless a new stimulus were to impose itself strongly and send the network on its way towards a new single attractor.

The extension of these dynamical processes to networks exhibiting ordered temporal sequences of 'cognitive events' implies that there can be no attractors, since a network which enters in a simple attractor remains in it indefinitely. However, since the presence in an attractor does not have to persist for longer than is required for the organism to become aware that an attractor has been reached, quasi-attractors are equally good, provided they last for a reasonably long period. But there are other requirements that the quasi-attractors should satisfy.

- They should allow for the storage and retrieval of a variety of temporal sequences in a *single* network.
- They should preserve the error correcting (associative recall, content addressing) nature of attractors.
- They should be robust against noise of various types.

This agenda was solved in the following way (Sompolinsky and Kanter 1986, Kleinfeld 1986). The synaptic organisation of the network was supplemented by a second set of synapses—slow synapses. Signals coming through such synapses would be delayed and consequently they would communicate information about past network states. This should be contrasted with the very short delay with which the first set was transmitting information. The other essential feature of the second set of synapses is that their values are such as to induce transitions between one attractor and another. In other words, the fast synapses define a whole variety of attractors. Each attractor is a network state. The efficacies of the slow synapses are designed or learned to bring about the transition of the network from one attractor state to another, based on the state of the network some time earlier. That time is determined by the delay inherent in the slow synapses. The various sequences of transitions are just the ordered temporal sequences which the network is capable of retrieving associatively.

The effectiveness of such networks is due to the fact that the fast synapses take care of error correction and robustness of the quasi-attractors, while the slow synapses start interfering only after a prolonged presence in one of the attractors of the sequence. This allows each pattern in the sequence to become a 'cognitive event'. Only after the network has stayed in one of the attractors of a sequence for a time longer than the delay of the slow synapses does the latter begin to transmit to the neurons the information that the network had indeed been present in that particular attractor. This is expressed in terms of afferent post-synaptic potentials (PSP) which act to change the distribution of neural activities so as to move the network into the next attractor state in the chain. These PSPs compete with the stabilising PSPs, arriving via the fast synapses. The course of the network then depends on the ratio of the amplitudes of the slow to the fast synapses. If it is large, a transition begins to take place. Once it starts it is reinforced by the fast synapses which tend to draw the network into the new state. The robustness and the attraction of the new state in the sequence are now enhanced by both types of synapses—the fast ones due to standard instant attraction and the slow ones through the continued memory of the presence in the previous attractor which continues to push in the same direction.

On the other hand, if the ratio is small, then the quasi-attractors becomes real attractors and the network gets stuck. The transitions can then be assisted by external inputs. The basic observation here is that this input does not have to be specific. In particular, if it is not associated with the stored patterns in the sequence, the same incoming distribution of afferent potentials can induce transitions down a temporal sequence which would not proceed spontaneously. The role of the transition (slow) synapses is to indicate the temporal order which can be stimulated. This feature has been employed to construct a network which can count identical stimuli (Amit 1988). It provides for a sequence of 'cognitive' states which can be reached conditionally and can be left conditionally. As we shall see in subsection 7.2 below this feature will be very helpful in designing the logical part of our network, whose task is to make logical decisions based on such predicates as the end of the rehearsed sequence or the occurrence of recognition in the process of rehearsal.

## 7. ANN recalling sequences and recognising probes

We now describe a network which, upon receiving a probe stimulus, recalls (rehearses) the sequence of the memory set, in the order in which it has been learned. Once initiated this rehearsal takes place spontaneously, in the robust and 'cognitive' manner described in section 6 above. While rehearsing the sequence a comparison with the probe is continuously performed. A response is generated only when the passage through the sequence has been completed. The response is positive or negative depending on whether during the rehearsal a match has, or has not, occurred between the probe and one of the items in the set.

### 7.1. *Modes of description*

The description of the network will necessarily take place in a few modes. First there is the structural versus the functional. The first is about the network and its network components, about mutual connections of networks and of neurons within networks. The second is about what each component does and how it influences the others, down to what neurons do in each component. Both modes divide further into the biological–physical and into the logical–behavioural. The first subdivision is expressed in the language of neurons, potentials, attractors and transitions. The second one is of cognitive recall, of recognition, of decision and of reaction.

Still another descriptive tool is the separation of the network into parts. We will deal with a single network in the sense that no subgroup of its elements is disconnected from the rest. Yet it can be meaningfully analysed in terms of three sub-networks. This is because some parts are connected internally with full feed-back, while between parts the connections are essentially one way, or feed-forward. Thus one can usefully discuss the operation of each sub-network as a separate ANN, including the effects of other sub-networks as external afferent influences.



### 7.2. *Functional structure*

*ANN-1: the probe (stimulus) buffer.* The function of the first sub-network, ANN-1 in figure 1, is to store the probe (stimulus) during the search (trial)†.

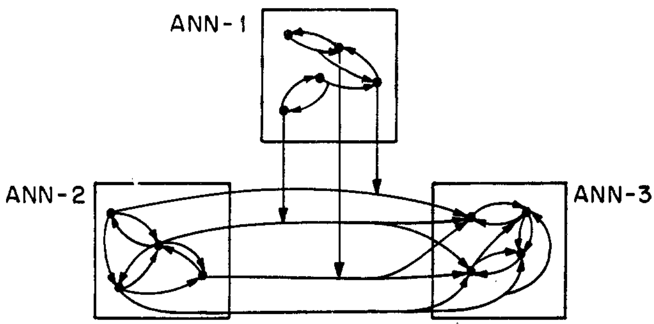


**Figure 1.** A three-component neural network performing the Sternberg experiment. Within each sub-network neurons are represented by black dots and arrows represent synaptic directionality. Sub-networks are feed-back systems, while inter-network communication is feed-forward.

ANN-1 is introduced to account for the apparent fact that performance is independent of the probe presentation time. This can be achieved in various ways. For example, it could make use of an attractor in long-term memory, if the probes are familiar objects. Or, alternatively, one can assume that the probe dynamically creates an attractor that stabilises a network state corresponding to the stimulus, once the

† We shall consider the pairs stimulus–probe and search–trial as pairs of synonyms. The first partner in each pair is more current in the psychocognitive domain, while the second has become more common in the context of network analysis.

minimum had been dug by the persistent stimulus. This may point to a distinction between experiments with memory sets of familiar objects and with random patterns. The latter would require longer presentation times for the probes. There is a third alternative, namely that the stimulus sensory data are present for the full length of the search (trial). In that case this network is redundant and the probe's sensory input can directly play its role in the integrated network. See the discussion in subsection 9.1.

ANN-2: *principal network storing memory sequence set.* The function of the second sub-network is ANN-2, in figure 1, is to rehearse the ordered memory set every time a probe is presented for trial. The memory set is encoded into this network, during the presentation (training) period, as a temporal sequence. It is a temporal sequence with strong transitions, as described in section 6. The attractors are quasi-attractors and the transitions down the sequence take place spontaneously, once the rehearsal is provoked by the probe. This network 'reverberates' in each of the patterns of the memory set for the duration of the synaptic delay, which is a period long enough for cognitive recall of each these patterns. If patterns appear repeatedly in the memory set they will be gone through repeatedly as the sequence is rehearsed, in their proper order. See figure 2.

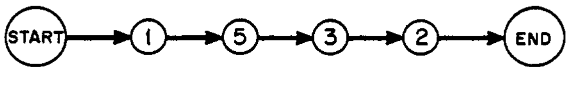


**Figure 2.** Structure of relevant attractors and of the transitions between them, as implied by the synaptic matrix of ANN-2. Circles are attractor network states and are represented by a symbol. The actual neuronal network states are fixed random uncorrelated states. Thick arrows indicate delayed synaptic contributions which are strong enough to provoke an unaided transition between the two attractors they connect. The attractors represented by numbers are the learned memory set. START and END are functional attractors, independent of the memory set.

When the sequence has been gone through, the network goes into one additional special state—the END state. This is a fixed pattern which indicates that the rehearsal has been completed. Clearly, the time for a complete rehearsal grows linearly with the number of patterns in the memory set. In the particular implementation we have chosen, following a complete passage through the temporal sequence, rehearsal is restarted. One could have just as well arranged for the network to move to some other state, foreign to the sequence, after the END state has been visited. We shall not be specific about the behaviour of the network following a complete passage through the sequence, which terminates in a response. At this point the probe is removed and attention is diverted to other tasks, until another probe is presented. In this sense the complete process involves a single repetition of the sequence and this unit applies equally well to a situation of constant memory set and to experiments with varied memory sets.

ANN-3: *the logical, decision-making network.* The last sub-network, ANN-3 in figure 1, is a 'logical' network. It encodes a universal, fixed (in the sense that it does not have to be specifically learned) weak temporal sequence of attractors, each of which is a long-lived state in the absence of external input, as described at the end of section 6. On the arrival of a probe this network enters the state marked A in figure 3. When ANN-2 goes through a pattern in the sequence which is similar enough to the probe, as present in ANN-1, it produces in ANN-3 an input which provokes a transition into the state marked B. This, again is a persistent cognitive state indicating that a recognition had occurred. If a second recognition takes place, which implies that some pattern in the memory set has been repeated, the logical network ANN-3 will make a transition into state C.

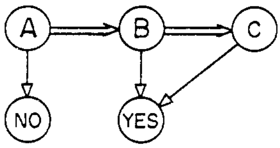


**Figure 3.** Attractors in the logical network ANN-3. All attractor states are functional and unrelated to the memory set. State A is where ANN-3 starts when a probe is presented in a new trial. B is an attractor implying recognition; C implies that recognition has occurred more than once. YES and NO are final decision states, which determine the reaction of the entire network to a probe. All arrows represent weak transition synapses. Double arrows are transition synapses stronger than single arrows.



Note that the arrows connecting states in figure 3 stand for weak delayed transition synapses which give a direction to transitions which can take place only if stimulated by the input arriving from ANN-2. All transitions in this network are weak. The single and double arrows in figure 3 indicate that the latter, while too weak to cause a transition, are stronger than the former. Thus the delayed synapses signalling transitions from A to B, from B to C and from C to YES are stronger than the ones signalling transitions from A to NO or from B to YES. The functional motivation for these synaptic differences is that ANN-3 should make a transition from A to B, or from B to C, on *every* recognition, and should avoid moving by mistake into one of the terminal states. The move into the latter (YES and NO) is helped by a specialised input communicated by ANN-2 when it reaches the END state. See subsection 8.1 below.

Finally, ANN-2 arrives at the END state. When ANN-3 receives information about this fact it makes a transition to state YES if it has been in one of the states B, C. If, instead, it has been in state A, which implies that no recognition had taken place, based on the same input from ANN-2, it will land in state NO. When ANN-3 has been in either state YES or state NO for a long enough duration the system responds with the corresponding reaction.

## 8. Neuronal and synaptic structure

As previously mentioned each sub-network is composed of neurons extensively connected, with full feed-back. Extensive connectivity signifies that each neuron is connected, both afferently and efferently, to a number of other neurons which is of the same order of magnitude as the total number of neurons in the network. Neurons in different sub-networks, when connected, are connected only one way (feed-forward). This is depicted schematically by the arrows in figure 1.

### *8.1. Synaptic connectivity*

*Inter-network connections*. In the present scheme there are two types of inter-network connections.

- Neurons in ANN-1 synapse on about one half of the neurons of ANN-2. One neuron of ANN-1 synapses on a single neuron of ANN-2. It does so by an inhibitory synapse whose role it is to block the spikes of the inter-network axon communicating from ANN-2 to ANN-3. As will be discussed below, this is our tool for recognition.
- Every neuron in ANN-2 is connected synaptically to every neuron in ANN-3, via axonic synapses. This is a standard kind of connectivity, excitatory and inhibitory, considered in the context of ANNs, except that it is unidirectional. The synapses from the neurons which are modulated by the probe have a uniform strength, uncorrelated with the attractors of ANN-3. The strength of the other synapses is correlated on the one hand with the END state in ANN-2 and with the YES and NO states in ANN-3 on the other. This provides special input, correlated with the YES and the NO states, into the logical network upon completion of rehearsal (see below). Note that since the END pattern as well as the YES and NO patterns can be fixed so can this synaptic structure.
- Neurons in ANN-2 whose output is modulated by the probe do not feed back into the dynamics of ANN-2.

A schematic representation of such a synaptic arrangement is presented in figure 4. It should be pointed out, though, that this separation is just a matter of convenience. It is not necessary, since the internal dynamics of the network is extremely robust.

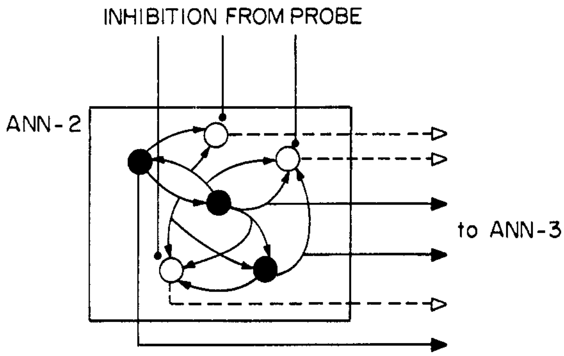


**Figure 4.** A schematic representation of the synaptic arrangement for affecting recognition. One half of the neurons in the box ANN-2 (open circles) have inhibitory afferent synapses, coming from the stimulus and blocking a potential outgoing spike, if the corresponding presynaptic neuron is active. These neurons do not participate in the feed-back of ANN-2, but receive the usual inputs. They send efferent axons (broken lines) to ANN-3. The rest of the neurons in ANN-2 (full circles) operate with full feed-back, and are connected to ANN-3 (full lines) by synapses correlated with END in ANN-2 and with YES and NO in ANN-3. See, e.g., subsection 11.1.

The essential feature is that usually one single neuron acting in an excitatory fashion on another will have a marginal effect. In order to be able to compare attractors one must have such one-to-one connections (see also subsection 11.3 below). This is envisaged to be the role of inhibitory synapses, whose strong nonlinear effect can block a whole spike by a single synapse. The strong effect is related to the fact that inhibitory synapses change the membrane's gain (Fatt and Katz 1953, Rall and Segev 1987) in contrast to excitatory synapses, each of which contributes a marginal additive term toward the generation of a spike. How this blocking solves the problem of recognition is explained in subsection 11.3. This is the role of one half of the neurons of ANN-2.

*Connectivity within ANN-1.* ANN-1 is a network of $N_1$ neurons which, for our purposes here, has only fast stabilising synapses to make the probe pattern into an attractor. When the probe is presented it provokes synaptic modifications to set up a synaptic matrix which stabilises this pattern itself. Our model will not explicitly deal with this aspect, since this part is self-evident.

*Connectivity within ANN-2.* The connectivity in ANN-2 consists of one set of fast synapses and one set of slow ones. The efficacies of the fast synapses have one fixed part, which does not have to vary from experiment to experiment. This is the contribution which stabilises the initial state START and the last state END in figure 2. When a memory set of $p$ patterns is taught to the network the fast synapses are modified to stabilise each of the $p$ patterns. In the present realisation all $p+2$ patterns, when coded by neural activities, are uncorrelated. In addition, the matrix of slow, delay synapses which induce transitions is formed. Transitions have to be induced from START to the first pattern in the sequence; from pattern to pattern in the sequence and finally from the last pattern in the sequence to state END. These two sets of synapses have the forms specified in subsection 11.1, namely all transition synapses are strong enough to produce transitions. The process by which they are imprinted is avoided again.

To allow for sequences with repetitions the mechanism of synaptic delays must be somewhat extended. Repetitions are included by allowing for synapses with additional delays. To produce sequences with patterns appearing at most twice we introduce one additional set of synaptic delays which have a longer delay and are stronger in amplitude. The function of these additional delays is to induce transitions based on the network's memory of its presence in the last two attractors. Such a chain of attractors is shown in figure 5. That this actually solves the construction of temporal sequences with single repetitions and that it can be generalised to multiple repetitions is shown in subsection 11.2.

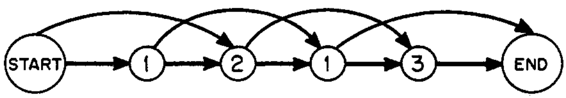


**Figure 5.** An example of the ANN-2 attractors for a memory set with repetitions. Long arrows are transition synapses with double delay.

Note also that since about one half of the neurons in ANN-2 are blocked by the spikes from ANN-1, the internal dynamics of this network will usually proceed with one half of the neurons driving the network among the attractors.

*Connectivity in ANN-3.* The connectivity in ANN-3 is fixed. There are fast synaptic connections which stabilise the attractors in figure 3. All contributions to the slow-transition synapses are weak so that transitions do not take place spontaneously. The various transitions included in the synaptic efficacies are marked in figure 3. The arrow from A to B is doubled to indicate that while none of these transitions will take place spontaneously, the transition toward B will be more likely. The same applies for the contributions to the slow synapses coming from the double arrow emanating from B, etc. We come back to this point in describing the dynamics below. The technical details are left to subsections 9.1 and 11.1.

### *8.2. Sub-network dynamics*

Given the presence of the probe stimulus, or the attractor which it had dug for itself, ANN-1 will persist in the probe state. Each of the neurons active in this pattern will send, at relatively high rate, efferent messages (spike bursts) to inhibit the transmission of the output of the corresponding set of neurons from ANN-2 to ANN-3, as mentioned above. The output of the corresponding neurons in ANN-2 is blocked throughout the trial. The blocking of the transmission to ANN-3 becomes particularly effective when the active neurons in ANN-2 are highly correlated with the probe. This indicates recognition.

When the probe is presented ANN-2 enters the attractor START and ANN-3 enters state A. ANN-2 remains in START for the duration of the shortest synaptic delay. Then it makes a transition into the first pattern of the memory sequence. The network continues down the sequence, remaining in every pattern for a duration of the short delay, and aided in ambiguous situations by the longer synaptic delays. The dynamical evolution of ANN-2 is independent of the events in the other sub-networks. The internal dynamics in the network operates with $+1$ and $-1$ representing spiking and refractory neurons, respectively.

At each step the *time-averaged* output of ANN-2 feeds, as post-synaptic potentials (PSPs), into ANN-3. The time averaging ensures that only attractors (cognitively significant events) in ANN-2 are communicated to ANN-3. Only firing $(+1)$ neurons send input into ANN-3. This input is modulated by the synaptic connectivity between the two sub-networks and by the stimulus which is present as a persistent output of ANN-1. The communication between these two networks is rather subtle:

- If ANN-2 reaches the END attractor, it will try to communicate to ANN-3 a PSP which is correlated with the YES and the NO states, due to the synaptic structure described above. This communication will not be disturbed by the inhibition coming from ANN-1 and blocking neurons which correspond to the probe, since this blocking has been chosen not to affect the internal dynamics. Note that even if we had not made this special choice there should have been no interference since the END state is uncorrelated with the stimulus. Thus, the PSP arriving in ANN-3 will be correlated with the two terminal states.
- If ANN-3 is in state A (no recognition) when this input arrives, it will move into the NO state, due to the correlation of the incoming PSP with this state.
- If ANN-3 is in states B, C, which indicates prior recognition, it will move into state YES, due to the correlation of that state with the arriving PSP.
- If ANN-2 is going through the sequence of attractor states which are uncorrelated with the probe, then there will be some typical average input into the neurons of ANN-3. This, unspecial input we assume to be cancelled against an approximately uniform threshold. Given this cancellation, ANN-3 will remain in whatever attractor it had been in, because the transition synapses are weak.
- When recognition takes place there is a significant reduction in the potential flow into the neurons of ANN-3. The surplus, which is uncorrelated with the attractors of this network, will provoke a transition, guided by the stronger delay synapses, i.e. from A to B, or from B to C, as in the chime-counting model (Amit 1988).
- If recognition has occurred, ANN-3 is in B. As long as there is not a second recognition nor an END attractor, the input into ANN-3 will not be sufficient to bring about another transition.
- A second recognition will drive ANN-3 from B into C, rather than into YES, by the same mechanism which drove it into B.

- Arrival of ANN-2 at END will transfer ANN-3 from B (or C) into YES, as it did from A into NO.

There is, of course, a difference between the status of B and C in figure 3. This difference expresses itself in a dynamical difference. The transition from C to YES on the arrival of ANN-2 to END is faster than from B to YES. The reason is that B has transition connections which act to assist transitions both to B and to YES. This causes ANN-3 to hesitate for a while before choosing YES. On the other hand, C is connected to YES only and the hesitation is spared. See also the discussion of the simulations in the next section.

## 9. Computer simulation of the operation of the network

### *9.1. Technical specifications*



In the simulation of the operation of the combined network we have actually implemented the coupled combined dynamics of ANN-2 and ANN-3 only. ANN-1 is represented by a fixed potential input (the probe) which modulates the communication between ANN-2 and ANN-3, as explained above. This is a simplified implementation of this sub-network but it introduces nothing artificial. Each of the two simulated networks was composed of 500 neurons. They could have been smaller, but this would be accompanied by an increased sensitivity to fluctuations in the correlations between the randomly chosen stored patterns. ANN-3 is particularly sensitive because of the delicate nature of some of the logical transitions that have to take place in it.

The basic synaptic delay time in both networks was chosen to be five neural cycle-times. In a neural cycle-time all neurons in the network update their activity states once. If there are repetitions in the memorised sequence, additional synaptic delays are brought in and they have been chosen as multiples of the basic synaptic delay (see, e.g., subsection 11.2). The consequences of variations in the relative delays in the two networks are briefly discussed in section 10. The averaging time in the transmission from ANN-2 to ANN-3 is four cycle-times. In other words, only attractor states are communicated across.

There is one additional, special, delay we introduce in communicating to ANN-3 the information that ANN-2 has reached the END attractor. Recall that on reaching this attractor ANN-2 communicates to ANN-3 a PSP which is proportional to a mixture of the YES and NO attractors in ANN-3. This is designed to deal with the special circumstance in which recognition takes place on the last item in the sequence. It can be avoided if one introduces an additional quasi-attractor between the last item and the END attractor.

The attractor states in each network were chosen as $N$-bit words of independent $+1$s and $-1$s, with equal probability. The synaptic efficacies within the two networks and between them are then constructed according to the prescriptions of subsection 11.1. Such states represent a spatial mean level of activity of 50%, which is rather unrealistic. This choice was made for convenience, without limiting the generality of the results. In fact, minor modifications of the storage prescription are known to provide for effective storage and retrieval of patterns of arbitrary levels of activity (Amit *et al* 1987b, Gardner 1988, Buhmann *et al* 1988, Tsodyks and Feigel'man 1988).

The dynamics is a sequential, asynchronous, stochastic (Glauber, heat-bath) process (Amit *et al* 1985), with a noise parameter (temperature) between 0.1 and 0.2. The amplitudes of the various synaptic coefficients, as defined in subsection 11.1, are as follows.

- The amplitudes of the fast stabilising terms is unity in both networks.
- The basic amplitude of the delayed synapses in ANN-2 is 1.2.
- The standard increase of consecutive delay amplitudes, in ANN-2, is 0.22. Thus the amplitude of the contribution of a double delay synaptic contribution would be 1.42.
- The transition amplitudes in ANN-3 are:
  Transition from A to B and from B to C: $b_1 = b_3 = 0.4$.
  Transition from A to NO and from B to YES: $b_2 = b_4 = 0.25$.
  Transition from C to YES: $b_5 = 0.7$.
- The inter-network synaptic amplitudes are:
  The amplitude of synapses communicating recognition: $g_1 = 1.4$. The effective strength is 0.35, because typically only one quarter of these synapses will be activated, given that the blocking affects one half of the axons and only one half of those participate in the pattern.
  The amplitude communicating the YES and NO input on arrival of ANN-2 at END: $g_2 = 1.05$.

The amplitude of the noise should be compared to a typical local PSP arriving at a neuron which is between 0.6 and 2.2, when ANN-2 is in a quasi-attractor and there is a single delay in action (single repetition), or between 0.6 and 3.6 if there are two delays. On the other hand, when ANN-3 is making its most delicate transitions, the PSPs can be as small as 0.1. But, it should be recalled that noise assists transitions and so this high relative noise causes no problems.

### 9.2. *Performance*

Below we shall follow the performance of the integrated network by displaying the overlaps of the current states of ANN-2 and ANN-3 with all the attractors of the corresponding network at different instances of time measured in updating cycles. Sample displays are shown in figures 6–8. In those figures the overlaps are represented by bars whose height is proportional to the corresponding overlap. If this value is close to unity, then every neuron in a subset corresponding to the pattern is emitting bursts of spikes, for the duration for which the network remains in the quasi-attractor, while all other neurons in the network are quiescent. The item (symbol) corresponding to the pattern represented by each bar is indicated under the bar. In each figure we also present the items in the memory set as well as the pattern constituting the probe.

When the simulation is initiated the stimuli are selected by the operator. The corresponding random patterns are then generated by the program. The number of delays is also prescribed by the operator and should correspond to the configuration of repeated items in the memory set, as explained in subsection 11.2. This replaces the process of learning the memory set. Then a probe is chosen and the operation starts. In all three figures one notes that ANN-2 moves immediately into attractor START and ANN-3 into A. Five cycle-times later ANN-2 is seen to move, in all cases, to an attractor with a full overlap with the first item in the memory set. During intermediate times, which are not exhibited in the figures, the networks remain in the quasi-attractors, reproducing the state of activity most recently displayed. From here on the behaviour of the network depends on the stimulus.

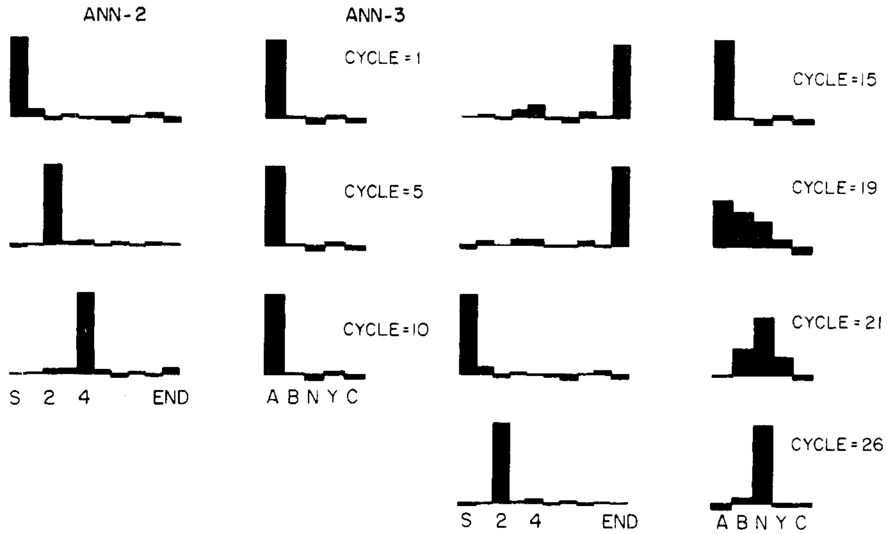


**Figure 6.** Negative probe: the memory set is 2, 4 and the probe is 7. Simulation with 500 neurons in ANN-2 and in ANN-3. Each pair of histograms shows the similarity of the state of the corresponding sub-network (left ANN-2; right ANN-3) to the memorised pattern indicated under the bar at the bottom of the figure, at the time marked at the top right of each display. Between cycle 1 and cycle 5 the situation is as in 1, implying that neurons active in S and in A are emitting bursts, and similarly for the periods between cycles 5 and 10, or between 10 and 15. The arrival of ANN-3 at an attractor with bursting neurons corresponding to pattern N (NO) is a signal that the system should respond negatively. Note that N (NO) and Y (YES) are stable attractors.



### *9.3. Negative probe*

Figure 6 represents the case in which the probe (stimulus) is not in the memory set. The memory set includes 2 and 4, while the probe is 7. One observes that while ANN-2 moves from one memory attractor to the next ANN-3 remains in A. After five cycles ANN-2 moves from START to state 2. It remains in it for five cycles, each neuron in the corresponding subset emits a burst of five spikes. Then, by cycle 10, ANN-2 moves into pattern 4. It remains there again for five cycles, a different subset of neurons emitting five spikes. ANN-3 remains in state A until cycle 19, despite the fact that ANN-2 has moved into END by cycle 15. This is due to the fact that ANN-2 communicates to ANN-3 only facts about attractors, and it takes time to establish that the state END, into which the ANN-2 has entered, is an attractor. Technically, this is expressed by the fact that the communication between the two networks averages the activities of the neurons in ANN-2.

By cycle 19 ANN-3 moves into an intermediate state which is a mixture of YES, NO and B, due to the input injected by ANN-2 while it is in the END attractor. In the mixture the NO state is more prominent due to the synaptic structure. Then, as ANN-2 moves out of the END state, and into START, ANN-3 begins to sort out its final decision, which is concluded somewhere between cycles 21 and 26. The conclusion is reached, in the present paradigm, when a sufficiently large fraction of the neurons, corresponding to pattern NO (or alternatively YES), emit bursts of spikes. The precise fraction required is determined, of course, by the read-out mechanism. But ANN-3 cannot do better than at cycle 26, where essentially every neuron corresponding to NO is firing bursts.

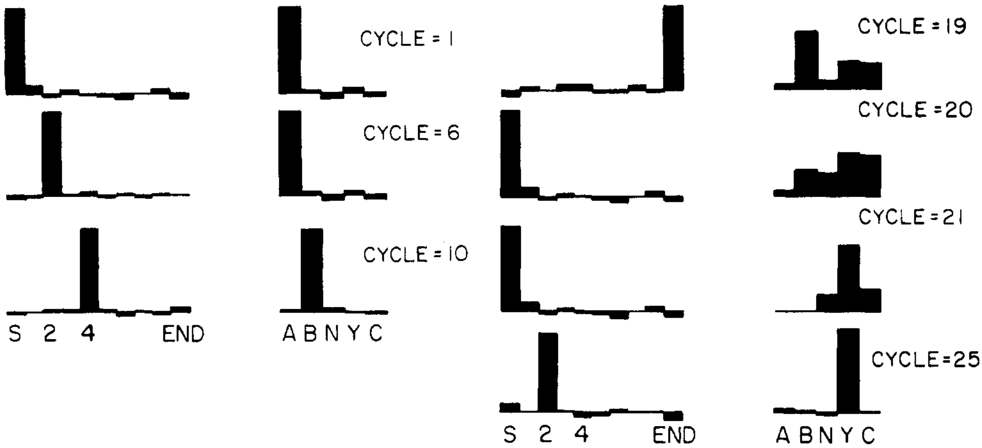


**Figure 7.** Positive probe: the memory set is 2, 4 and the probe is 2. The arrival of ANN-3 at an attractor with bursting neurons corresponding to pattern Y is a signal that the system should respond positively. For details, see the caption of figure 6.



### *9.4. Simple positive probe*

This case is presented in figure 7 and represents the experiment in which the probe appears once in a memory set with no repetitions. The memory set is as in the previous experiment. The probe is the symbol 2. ANN-2 begins in START and by cycle 6 (figure 7) it is fully in the state 2. Yet, the transition of ANN-3 from A to B, corresponding to the fact that recognition has taken place, occurs only by cycle 10. The additional delay is required, again, in order to establish that the state 2 in ANN-2 is indeed an attractor.

Note that by cycle 10 ANN-2 enters state 4 and remains there until cycle 15, at which time it moves into state END. But only by cycle 19 does ANN-3 learn about this fact and start to move out of B, to make its final decision. First it enters a mixture of states, in cycle 20, and when the influence of the mixed input due to END is removed ANN-3 starts to move clearly into attractor YES, by cycle 21. The process is cleaned up by cycle 25.

But recognition is not always that easy, as illustrated in figure 8. The situation is essentially the same as in the previous experiment, i.e. the same memory set and the same probe. The random numbers, though, are different. Until cycle 16 both experiments lead to very similar internal dynamics. But then cycles 21–29 show that ANN-3 has a difficult time deciding between the YES and the NO. It finally settles on the right answer, but the time is significantly longer.

### *9.5. The case of repetitions*

We do not present simulations for the performance of the network when there are repetitions in the memory set and the probe is not in the set. The reason is that in this case the network automatically performs as if the repeated item were in the set a multiple number of times. ANN-2 would go through the full memory set, aided by the additional delay synapses, while ANN-3 remains waiting in A, until ANN-2 has completed its tour.

With a positive probe the situation is quite different. This is exemplified in figure 9. The memory set is 2, 4 and 2 and the probe is 2. By cycle 5 ANN-2 goes into the attractor 2, and some time later, at cycle 8, ANN-3 starts to move into B. This is the first

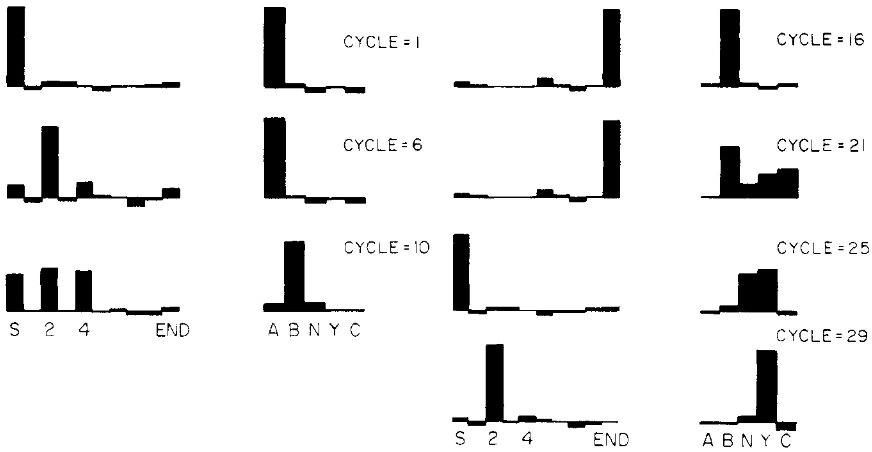


**Figure 8.** Positive probe, RT fluctuation: same as the experiment in figure 7 (memory set 2, 4 and probe 2) except that the random numbers used in the simulation are different. The result is a correct answer in a longer time. For details, see the caption of figure 6.

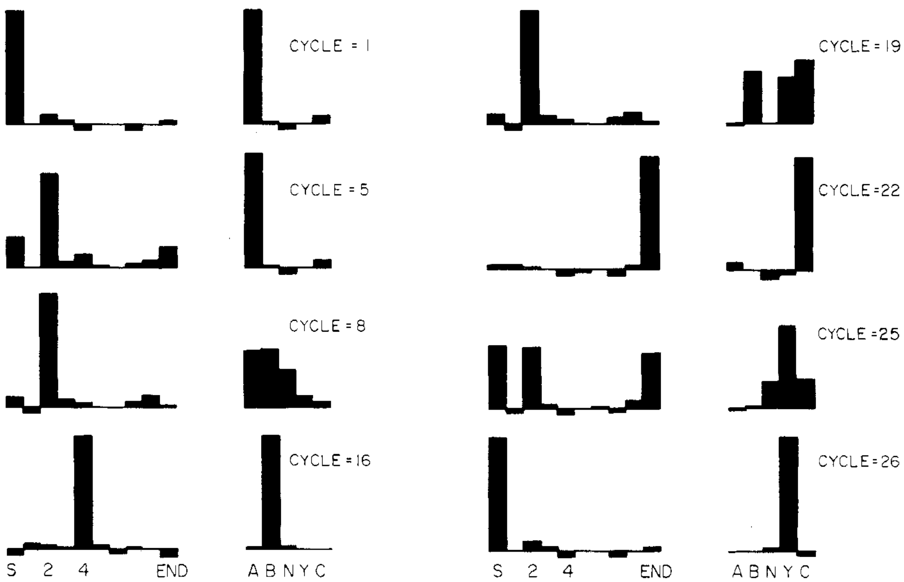


**Figure 9.** Positive probe with repetition: the memory set is 2, 4, 2 and the probe is 2. Note that following the first recognition, ANN-3 is in attractor B until cycle 16; it goes into attractor C upon the second recognition. It then goes rapidly into Y. See also, e.g., section 10. The memory sequence is activated with two synaptic delays. For details, see the caption of figure 6.

recognition. ANN-2 then moves to 4 and back to 2. The presence of ANN-3 in B together with a second recognition, provokes by cycle 19 the beginning of a transition of ANN-3 into C, which is an attractor signifying double recognition. The transition into C is completed by cycle 22 and ANN-3 finds it very easy to hop into the YES conclusion, once ANN-2 has moved through the END state. It does not hesitate that long in mixture states. The RT is 26 cycles, a result that captures nicely the speed-up associated with the recognition of positive probes which appear repeatedly in the learning set. The reason for the easier final transition is that while B is connected to both YES and C, C is connected to YES alone. This saves a long hesitation in mixed intermediate states. See, e.g., the discussion at the end of subsection 8.2.

### *9.6. Sample statistics*

For illustration purposes the network has been run ten times for each of three types of experiments, each with three sizes of the memory set. The reaction times (RT) are presented in table 1. In each entry in the table we present the mean of the reaction times over the ten trials as well as the RMS deviation, in parentheses.

**Table 1.** Sample statistics of network reaction times for several tasks as a function of the number of items in memory set, with RMS deviations in parentheses.

| Number of items | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| RT positive | 20.9 (0.7) | 25.4 (0.5) | 29.7 (2.0) | 33.7 (3.0) | 39.8 (2.0) |
| RT negative | 21.0 (0.4) | 25.5 (2.4) | 31.0 (1.0) | 36.0 (1.0) | 41.0 (1.0) |
| RT positive 1 repetition | — | 22.9 (2.0) | 25.6 (1.0) | 29.5 (3.5) | 32.0 (5.0) |

Note that the RMS fluctuation in the RT for the recognition with repetition is typically larger. This is a result of the fact that, for most of the trials, the answer is given when ANN-3 receives the input signalling that ANN-2 has reached the END state. However, in some cases the transition to attractor C occurs before the second recognition (due to random fluctuations related to random correlations between the patterns) and the second recognition initiates the YES response. The same phenomenon can, of course, also take place when there is no repetition, since the C attractor is always there. Yet, in that case there is no significant fluctuation in the RT because the final transition must still wait for ANN-2 to arrive at the END attractor.

## 10. Discussion

The network presented in the preceding sections has been designed to reproduce the gross features of the systematic behaviour in fast recognition experiments. The main challenge in this engineering endeavour has been to contend with methodological and with biological constraints. This is where engineering becomes an intellectual challenge. The main methodological constraint has been the insistence that all cognitively meaningful stages be represented by attractors, which in turn express themselves as selective bursting. The main biological constraint is related to recognition, or to the comparison of attractors. This, we have insisted, cannot be done either by the introduction of a vast number of triple synapses, or by giving single excitatory synapses a major influence on the behaviour of the post-synaptic neuron. We had to resort to the inhibitory mechanism, which stands a chance.

Once the network has been designed it acquires a life of its own. Questions can be asked about the effects on 'behaviour' of various modifications and 'pathologies' can be observed. Those, in turn, can suggest psychophysical experiments and/or modifications of the network. We now turn to a discussion of a sample of such options.

### *10.1. Variations in relative synaptic delays*

The network has been designed to have equal synaptic delays in ANN-2 and in ANN-3. This was done in order to economise on the number of free parameters. It is natural to enquire about the effects of variations in the relative magnitudes of these delays. The operation of the network is robust under small variations of these delays, but some noticeable modifications take place when the variations grow larger.

*Longer delays in ANN-2.* Denoting the delay in ANN-2 by $\tau_2$, and that in ANN-3, by $\tau_3$, we consider first the case when $\tau_2 \gg \tau_3$. In that case, the main effect is that the recognition signal coming into ANN-3 from ANN-2 may be long enough to provoke more than one transition in ANN-3. If $\tau_2 > 3\tau_3$, then the process will be self-terminating upon presentation of a positive probe. In other words, recognition will produce a positive response before the sequence in ANN-2 has been exhausted. If $3\tau_3 > \tau_2 > 2\tau_3$, then upon recognition ANN-2 will drive ANN-3 into state C, which produces a speed-up of the positive reaction just as in the case of repetition. If there is no C attractor in ANN-3 (see below) then this delay difference will produce a self-terminating positive response. In fact, a negative response may also be altered by the difference in the delays. When the ANN-3 has to make its final choice between YES and NO it goes first into a mixture state where it remains until the END signal from ANN-2 stops (see, e.g., figure 6). If the END signal remains on for too long, by the time it fades out ANN-3 no longer has the information that it was previously in A and an error is likely to occur.

It is interesting to note that the correlation of slow scanning with self-termination is consistent with experimental observation (Clifton and Birenbaum 1970, Sternberg 1975).

*Longer delays in ANN-3.* The response may be mistaken in two special cases. The first is if recognition takes place on the first item in the sequence, since the recognition signal from ANN-2 will disappear before the transition from A to B is activated. The second case would be if recognition takes place on the last item in the sequence. The reason is that by the time the END information arrives and goes away, ANN-3 has not had the transition from B to YES activated.

Both cases lead quite naturally to possible experimental tests, whereby special fluctuations in the error rates for the deviating cases could be looked for.

### *10.2. In the absence of a C attractor*

The C attractor in ANN-3 was an *ad hoc* construct introduced to provide a special pathway for positive probes which are repeated in the memory set. It is natural to ask what would be the behaviour of the network in the absence of this device. We have investigated this question by simulations. It is found that if the values of all the other parameters are left unchanged, then there is no speed-up on recognising a repeated item. On the other hand, if the parameter $b_4$ in ANN-3, (i.e. the strength of the YES component in the transmission from ANN-2 upon arrival at END; see subsection 11.1) is increased to 0.5, the second recognition brings about the reaction, i.e. self-termination. The behaviour in all other cases remains unchanged. Self-termination is clearly a speed-up, which may be quantitatively wrong. But one must keep in mind that there is not much room for experimental variation of the distance of the second member of a pair of repeated items from the end of a sequence which is never longer than six items.

## 11. Technicalities

### *11.1. Synaptic structures*

Each of the networks can have an independent number of neurons as long as these numbers are moderately large, to decrease the effect of fluctuations. Some of these fluctuation effects can be reduced by an appeal to more elaborate synaptic structures such as the pseudo-inverse (Kohonen and Rouhonen 1973, Personnaz *et al* 1985) or to a learning algorithm (Gardner 1988). We prefer to construct the network using the explicit, simple Hopfield forms.

*ANN-1*. In this network we exhibit only the stabilisation of the probe. It is represented by a word of $N_1$ random $\pm 1$s, selected anew on every trial. To stabilise it in ANN-1 we will choose

$$J_{ij}^1 = \xi_i^{\text{PROBE}} \xi_j^{\text{PROBE}}.$$

*ANN-2*. This network will have $N_2$ neurons. To store a memory set of $p$ patterns it will have $p+2$ quasi-attractors† stabilised by fast synapses. These again will be random words of $\pm 1$s. They are denoted as follows:

$\xi^{\text{START}}$ START attractor

$\xi^{\mu}$ items in memory set $(\mu = 1, \ldots, p)$

$\xi^{\text{END}}$ END attractor.

The fast synaptic matrix is

$$J_{ij}^2 = \xi_i^{\text{START}} \xi_j^{\text{START}} + \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu} + \xi_i^{\text{END}} \xi_j^{\text{END}}. \tag{1}$$

The transition, single-delay, synapses will have the form:

$$T_{ij}^2(\tau) = \lambda_1 \sum_{\mu=0}^{p} \xi_i^{\mu+1} \xi_j^{\mu}. \tag{2}$$

Note that we have included the start pattern as number 0 and the END pattern as number $p+1$ in the above sum. The single delay is indicated by the argument $\tau$ of $T$.

*ANN-3*. In this network we have five attractors: A, B, C, YES and NO. They will be denoted by $\chi^{\mu}$, $\mu = 0, \ldots, 4$, respectively. The fast synaptic matrix is

$$J_{ij}^3 = \sum_{\mu=0}^{4} \chi_i^{\mu} \chi_j^{\mu}. \tag{3}$$

† This number may increase to $p+3$, for convenience. See, e.g., subsection 9.1.

The transition matrix will have the form:

$$T_{ij}^3(\tau) = b_1\chi^A\chi^B + b_2\chi^A\chi^{NO} + b_3\chi^B\chi^C + b_4\chi^B\chi^{YES} + b_5\chi^C\chi^{YES}.$$

*ANN-2 to ANN-3*. The synaptic matrix connecting ANN-2 to ANN-3 reads as follows. One component, the one that can be blocked by the inhibiting stimulus, is uniform, i.e. independent of the structure of the attractors in either network, and of magnitude $g_1$. In other words, every active neuron in this group of ANN-2 contributes approximately equally to every neuron in ANN-3. Recall that the activity is averaged before it is communicated.

The rest of the neurons in ANN-2 are connected as follows:

$$J_{ij}^{2,3} = g_2\xi_i^E(\chi_j^{YES} + \chi_j^{NO}). \tag{4}$$

*11.2. Moving across a sequence with repetitions*

To allow for repetitions in a temporal sequence of quasi-attractors, synapses with more than one time delay have to be introduced. For example, if ANN-2 is to follow the sequence of attractors given by:

$$\xi^{START} \to 1 \to 2 \to 1 \to 3 \to \xi^{END} \tag{5}$$

where $\xi^{START}$ and $\xi^{END}$ represent the START and END attractors, and 1, 2, etc, stand for the memory set $\xi^1, \xi^2 \ldots$. When the network reaches attractor 1, then following a synaptic delay time, this attractor will be destabilised, yet the network does not have sufficient information on whether to proceed to state 2 or to state 3. It would need information on at least one previous state.

A synaptic structure which would resolve this ambiguity is:

$$J_{ij}^1 = \frac{1}{N}\sum_\mu \xi_i^\mu\xi_j^\mu \qquad \text{(synapses without delay)} \tag{6}$$

$$J_{ij}^2 = \frac{1}{N}\lambda_1\sum_\mu \xi_i^{\mu+1}\xi_j^\mu \qquad \text{(synapses with delay } \tau) \tag{7}$$

$$J_{ij}^3 = \frac{1}{N}\lambda_2\sum_\mu \xi_i^{\mu+2}\xi_j^\mu \qquad \text{(synapses with delay } 2\tau). \tag{8}$$

In order to follow sequences of attractors with more repetitions, we will also need more time delays. Here there are synapses with one delay as well as synapses with a longer delay, which for simplicity was chosen to be twice as long.

First we proceed to show that the synaptic structure which is the sum of (6)–(8) can lead the network through the sequence of attractors (5). (An exception will be a sequence in which the repeated items are consecutive. This case will be discussed later.) The potential accumulated on each of the neurons after the net has stayed, for the first time around, in attractor 1 for a time $\tau$ is:

$$h_i(t) = \sum_{j=1}^N J_{ij}^1 S_j(t) + \sum_{j=1^N} J_{ij}^2 S_j(t-\tau) + \sum_{j=1}^N J_{ij}^3 S_j(t-2\tau) \tag{9}$$

where $S_i(t-\tau) = \xi_i^1$, and $S_i(t-2\tau) = \xi_i^{\mathrm{START}}$, and therefore

$$\begin{aligned} h_i &= \xi_i^1 + \lambda_1(\xi_i^2 + \xi_i^3) + \lambda_2\xi_i^2 \\ &= \xi_i^1 + \xi_i^2(\lambda_1 + \lambda_2) + \xi_i^3\lambda_1. \end{aligned} \tag{10}$$

If the second delayed synapses were not present, i.e. if $\lambda_2 = 0$, then the network would be unable to 'decide' into which pattern to proceed, and it would stay in a mixture of attractors 1, 2 and 3. In our case, choosing both $\lambda_1$ and $\lambda_2 > 1$, the coefficient of attractor 2 prevails and the network undergoes a transition to attactor 2. Immediately after the transition the state 2 is very stable. The neural potentials are:

$$\begin{aligned} h_i &= \xi_i^2 + \xi_i^2(\lambda_1 + \lambda_2) + \xi_i^3\lambda_1 \\ &= \xi_i^2(\lambda_1 + \lambda_2 + 1) + \xi_i^3\lambda_1 \end{aligned} \tag{11}$$

where we have used the fact that the delayed potentials remain unchanged immediately after the transition, for a time of $\tau$ cycles.

Following another period $\tau$, the potentials change to

$$h_i = \xi_i^2 + \xi_i^1(\lambda_1 + \lambda_2) + \xi_i^{\mathrm{END}}\lambda_2.$$

Note that a term proportional to $\xi^{\mathrm{END}}$ has appeared. It is due to the second delayed connection from attractor 1 to the attractor $\xi^{\mathrm{END}}$. Since both $\lambda_1$ and $\lambda_2 > 1$, the network will move toward 1. After the next delay time, $\tau$, the potentials become

$$h_i = \xi_i^1 + \xi_i^3(\lambda_1 + \lambda_2) + \xi_i^2\lambda_1$$

and this time the transition will proceed from 1 to 3. The difference is that the information about the past states, arriving via the delayed synapses, is not the same as it was when the network visited attractor 1 for the first time. This time around $S(t-\tau) = \xi^1$ and $S(t-2\tau) = \xi^2$.

If the repeated items are consecutive in the sequence, as, for example, in

$$\xi^{\mathrm{START}} \to 1 \to 2 \to 2 \to 3 \to \xi^{\mathrm{END}} \tag{12}$$

then two sets of delayed synapses, as in the previous scheme, would not suffice. The problem arises because the repeated attractor, 2 in the example, is connected by synapses of delay $\tau$ to itself and to attractor 3 and by synapses of delay $2\tau$ to 3 and to END. Consequently, after the network stays for a time $2\tau$ in the repeated attractor, 2, the potentials will be

$$h_i = (1 + \lambda_1)\xi_i^2 + (\lambda_1 + \lambda_2)\xi_i^3 + \lambda_2\xi_i^{\mathrm{END}}$$

which *does not attract to* any of the memorised attractors.

In this case the existence of a set of synapses with yet a longer delay, connecting attractor 1 to 3, will help the network choose the right course. One can also have a solution with double delays alone, if one allows for selective transition synapses. Instead of connecting *every* attractor in the sequence, by double-delay transition synapses, to an attractor two positions down the sequence, one could introduce double-delayed synapses only for the transition between 2 and 3. Leaving the single delay synapses as before. In this case, the term with $\xi^{\mathrm{END}}$ in the potential is absent and the transition to 3 proceeds properly. Such an arrangement will not harm the operation of the network with a sequence of the type of equation (5), where one would introduce double-delay synapses only for the transitions between START and 2 and between 2 and 3. We will not discuss here the relative merits of the two solutions.

*11.3. Probe recognition and inhibition*

Here we present a possible mechanism by which an attractor arrived at in ANN-2, can be compared for similarity with the stimulus (the probe) which is constantly present. There are two basic constraints. One is that the effect of recognition, i.e. the fact that the attractor is indeed similar to the probe, will be independent of the particular structure of the two compared patterns. The other is that the demands on the neurobiological hardware should not be too excessive. The effect must be on the behaviour of ANN-3, which should receive significantly different afferent input upon recognition than in the absence of recognition, so as to allow it to make a transition between its own attractors in the first case and not to make it in the second case.

To allow for such a comparison there must be a correspondence between the neurons of ANN-1 and those of ANN-2, such that when the activity of corresponding neurons of ANN-1 and ANN-2 is strongly correlated there will be a special signal. One possible mechanism might be that the outputs of corresponding pairs of neurons in ANN-1 and of ANN-2 be summed when arriving in ANN-3. Then, coincident activity of neurons in the two networks will produce, pair by pair, a larger effect than different activity. This type of solution is excluded, we believe, by neurobiological constraints. The effect of a single excitatory input on the spiking activity of a cortical neuron is marginal. Hence, it is hard to expect a major effect in ANN-3. On the other hand, individual inhibition promises to provide a much stronger effect if placed near the soma of a neuron (Fatt and Katz 1953 , Rall and Segev 1987), due to its shunting action. This can be employed to construct the desired mechanism.



Suppose that $S_i^1$ and $S_i^2$ represent the states of ANN-1 and ANN-2, respectively. These variable are $\pm 1$s. It is more natural to describe the process in terms of (1,0) variables, which actually enumerate spikes, namely $V_i^1 = (S_i^1+1)/2$ and $V_i^2 = (S_i^2+1)/2$. Suppose next that each neuron in ANN-3 receives a uniform input from ANN-2, mediated by an excitatory synaptic matrix $K_j \geq 0$, such that the potential on neuron $i$ in ANN-3 is

$$h_i^3 = \frac{1}{N_2}\sum_{j=1}^{N_2} K_j V_j^2$$

which is essentially independent of the receiving neuron. $N_2$ is the number of neurons in ANN-2. If $K_j$ does not fluctuate much, then $h_i$ will be essentially

$$h_i^3 \approx \langle K \rangle f$$

where $f$ is the fraction of neurons that would be active in a typical attractor. It would be 1/2 in the standard Hopfield model.

Next, the neurons of ANN-1 are arranged in such a way that each neuron in that network has an inhibitory synapse near the soma of the corresponding neuron in ANN-2. A spike in a neuron in ANN-1 can block a spike in the corresponding neuron in ANN-2. In this case the input of a neuron in ANN-3 will be

$$h_i^3 = \frac{1}{N}\sum_{j=1}^{N_2}(1 - V_j^1)V_j^2.$$

In other words, only those neurons active in ANN-2, whose counterparts in ANN-1 are inactive, will contribute. If the activity pattern in ANN-1 is uncorrelated with that of ANN-2, then typically

$$h_i^3 \approx \langle K \rangle f(1-f)$$

while if the two patterns are strongly correlated, i.e. recognition has occurred, then $h_i^3 \approx 0$. Hence, ANN-3 will get a significantly different input in the two cases. The particular input will depend only on whether recognition has or has not occurred, and not on the structure of the recognised pattern.

Finally, if the inputs are indeed roughly equal one can suppose that the value of the input for no recognition is approximately cancelled by a threshold equal to $\langle K\rangle f(1-f)$. Then the different input arriving upon recognition can provoke the desired transition corresponding to recognition.

## Acknowledgments

Most of this work was performed while DJA and DS were members of the Institute of Advanced Studies, at the Hebrew University in Jerusalem. This was an invaluable opportunity for interdisciplinary work. DJA's work was supported in part by a grant of the US–Israel Binational Science Foundation.



## References

Amit D J 1988 Neural networks counting chimes *Proc. Natl Acad. Sci. USA* **85** 2141

—— 1989 *Modeling Brain Function* (Cambridge: Cambridge University Press)

Amit D J, Gutfreund H and Sompolinsky H 1985 Spin-glass models of neural networks *Phys. Rev.* A **32** 1007

—— 1987a Statistical mechanics of neural networks near saturation *Ann. Phys., NY* **173** 30

—— 1987b Information storage in neural networks with low levels of activity *Phys. Rev.* A **35** 2293

Anderson J R and Bower G H 1974 *Human Associative Memory* (Washington, DC: Hemisphere)

Anderson R A and Mountcastle V B 1983 The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex *J. Neurosci.* **3** 532

Atkinson R C and Juola J F 1974 Search and decision processes in recognition memory *Contemporary Developments in Mathematical Psychology* vol 1, ed D H Krantz, R C Atkinson, R D Luce and P Suppes (San Francisco: Freeman)

Baddeley A D and Ecob J R 1973 Reaction time and short-term memory: implications of repetition effects for the high-speed exhaustive scan hypothesis *Q. J. Exp. Psychol.* **25** 229

Bloom F E, Lazerson A and Hofstadter L 1985 *Brain Mind and Behavior* (San Francisco: Freeman) p 191

Buhmann J, Divko R and Schulten K 1988 Associative memory with high information content *Preprint* Technische Universität München

Buhmann J and Schulten K 1987 Influence of noise on the function of a "physiological" neural network *Biol. Cybern.* **56** 313

Cavanagh J P 1972 Relation between the immediate memory span and the memory search rate *Psych. Rev.* **79** 525

Clifton C and Birenbaum S 1970 Effects of serial position and delay of probe in a memory scan task *J. Exp. Psychol.* **86** 69

Fatt P and Katz B 1953 The effect of inhibitory nerve impulses on a crustacean muscle fiber *J. Physiol. (Lond.)* **121** 374

Gardner E 1988 The phase space of interactions in neural network models *J. Phys. A: Math. Gen.* **21** 257

Hopfield J J 1982 Neural networks and physical systems with emergent selective computational abilities *Proc. Natl Acad. Sci. USA* **79** 2554

—— 1984 Collective processing and neural states *Modeling and Analysis in Biomedicine* ed C Nicollini (Singapore: World Scientific)

Kleinfeld D 1986 Sequential state generation by model neural networks *Proc. Natl Acad. Sci. USA* **83** 9469

Kohonen T and Rouhonen M 1973 Representation of associated data by matrix operators *IEEE Trans Comput.* **TC-22** 701

Kristofferson M W 1972a When item recognition and visual search functions are similar *Percept. Psychophys.* **12** 379

—— 1972b Effects of practice on character-classification performance *Can. J. Psychol.* **26** 540
Lively B L 1972 Speed/accuracy trade off and practice as determinants of stage durations in a memory search task *J. Exp. Psychol.* **96** 97
Lively B L and Sanford B J 1972 The use of category information in a memory search task *J. Exp. Psychol.* **93** 379
McNicol D and Stewart G W 1980 Reaction time and the study of memory *Reaction Times* ed A T Welford (New York: Academic)
Parisi G 1986 Asymmetric neural networks and the process of learning *J. Phys. A: Math. Gen.* **19** L675
Personnaz L, Guyon I and Dreyfus G 1985 Information storage and retrieval in spin-glass-like neural networks *J. Physique Lett.* **46** 359
Rall W and Segev I 1987 Functional pssibilities for synapses on dendrites and on dendritic spines *Synaptic Function* ed G Edelman *et al* (New York: Wiley)
Ross J 1970 Extended practice with a single-character classification task *Percept. Psychophys.* **8** 276
Shinomoto S 1987 A cognitive associative memory *Biol. Cybern.* **57** 197
Sompolinsky H 1986 Neural networks with non-linear synapses and static noise *Phys. Rev.* A **34** 2571
—— 1987 The theory of neural networks: The Hebb rule and beyond *Heidelberg Colloquium on Glassy Dynamics* ed L van Hemmen and I Morgenstern (Berlin: Springer)
Sompolinsky H and Kanter I 1986 Temporal association in asymmetric neural networks *Phys. Rev. Lett.* **57** 2861
Standing L 1973 Learning 10,000 pictures *Q. J. Exp. Psychol.* **25** 207
Sternberg S 1966 High-speed scanning in human memory *Science* **153** 652
—— 1969 Memory-scanning: mental processes revealed by reaction-time experiments *Am. Sci.* **57** 421
—— 1975 Memory scanning: new findings and current controversies *Q. J. Exp. Psychol.* **27** 1
Theios J, Smith P G, Haviland S E, Traupmann J and Moy M C 1973 Memory scanning as a serial self-terminating process *J. Exp. Psychol.* **97** 323–36
Townsend J T 1971 A note on the identifiability of parallel and serial processes *Percept. Psychophys.* **10** 161
Tsodyks M V and Feigel'man M V 1988 The enhanced storage capacity in neural networks with low activity level *Europhys. Lett.* **6** 101