Principle of Superposition-free Memory

MICHAEL CONRAD

*Department of Computer and Communication Sciences,
University of Michigan, Ann Arbor, Michigan* 48104, *U.S.A.*

The requirement that memory be superposition-free imposes a restrictive condition on the structure of neural models. A simple model which satisfies this condition consists of two levels of neurons, with primary level neurons firing in response to external input and secondary (reference) neurons modifying activated primaries in such a way that they can later fire these primaries. Reference neurons are either activated by other reference neurons (time-ordered memories) or by primaries (content-ordered memories). The model allows for general powers of memory manipulation (e.g. formation of associative memories) through remem-orization and has a number of implications which correspond to features of the brain.

A basic requirement for any effective memory system is that it be super-position-free. Imagine, for example, the difficulties which would accrue to a Turing machine in which symbols marked on the tape superimposed (or mixed) with one another. There is compelling evidence that biological systems do not suffer such difficulties, e.g. that if an organism is exposed to scene S and later to S', the memories of these two scenes remain pure (although it is possible that an imagined scene is constructed by recombining different parts of the original scene). As a simple and classical example, Pavlov found that in conditioning experiments circles and ellipses could always be distinguished until the latter assumed a definite, high degree of circularity (Pavlov, 1972).

In computer systems the basis of superposition-free memory is the addressable structure of memory. Each bistable memory element can be uniquely accessed and switched by activating the appropriate two lines. This simple concept of addressing cannot extend to neural networks, how-ever, for in general these have converging inputs and diverging outputs. A plausible and frequently made assumption is that in such networks particular memories are in general concomitant to patterns of neural firing— after all the original input (e.g. scene) in general activates some pattern

of firing. The problem of memory is thus to reconstruct this pattern, but in response to a secondary pattern of firing concomitant to an event other than the original input (for example, part of the original input or some other input with which it has been associated). Presumably this means that at least some neurons responding to the original, primary input (S_1) are modified by neurons firing in response to the secondary input (S_2). This is implicit in the frequently made assumption of cell assembly theory that contacts between neurons whose firing is temporally contiguous are strengthened (cf. Hebb, 1949). Now suppose that the organism is exposed to a second scene (S_1') and that the neurons firing in response to this scene are modified (as above) by neurons firing in response to a second secondary input (S_2'). *The condition for superposition-free memory is: no neuron whose firing is concomitant to both S_2 and S_2' modifies any neuron whose firing is concomitant to S_1 but not S_1' or to S_1' but not S_1.*

There are undoubtedly many solutions to the superposition problem, including solutions which attempt to approximate it away in the law of large numbers. In this note I want to describe a simple, direct solution, which appears to be fruitful in terms of its biological consequences.

The solution (to be called the reference neuron scheme) has five postulates (Figs 1 and 2).

(1) Certain neurons (to be called primaries) undergo sensitization at the dendrites (for example, by an antidromic wave of excitation or a local reverberation) when firing, e.g. in response to external inputs.

(2) The sensitized primaries are modified ("loaded") by other neurons (to be called reference neurons) and in such a way that they fire in response to the future firing of these reference neurons.

(3) Reference neurons competent to reconstruct a pattern of neural firing associated with one input are never used to load primaries sensitized as a consequence of another input.

(4) Reference neurons activated for loading primaries may themselves be loaded by primaries which contact them.

(5) The conditions under which reference neurons are active for loading primaries or calling primaries previously loaded by them is controlled by a supervisory network of neurons.

The storage part of the memory cycle thus consists of the firing and therefore sensitization of primaries (for example, by exposure to an external input) along with loading of sensitized primaries by reference neurons firing at about the same time or immediately thereafter. Loading of a primary by a reference neuron (or a reference neuron by a primary) amounts to opening up a dendritic-axonic contact. Thus the future firing of this reference neuron automatically reconstructs the pattern of activity associated with the original
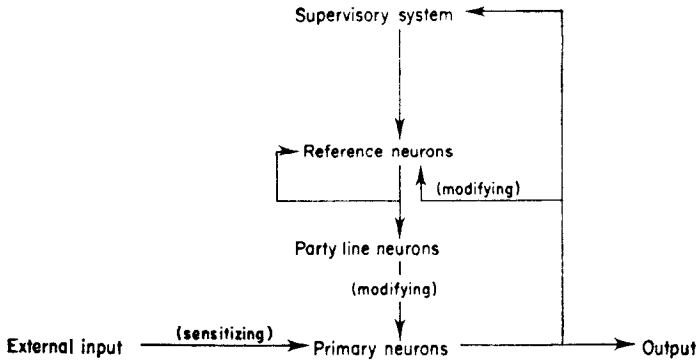
FIG. 1. Flow diagram of memory cycle in reference neuron scheme. External inputs (from preprocessing neurons) sensitize primaries in such a way that they can be modified (or loaded) by any reference neurons firing at the same time. Any such reference neuron is then capable of refiring (calling) the primaries loaded under its control, thereby reconstructing a pattern similar to the one induced by the original input. Reference neurons may also be loaded by primaries, allowing for content-ordered and associative memory structures, or they may be activated by one another, allowing for time-ordered memory. The supervisory system determines whether reference neurons are activated for loading or calling and also influences which reference neuron is activated (e.g. one receiving highest primary activation, one receiving activation from other reference neurons). Such supervision is mediable by inhibitory inputs to the reference neurons and is presumably influenced by inputs from the primaries and other parts of the brain. The party line neurons are not logically necessary but greatly increase the number of reference to primary contacts.

input, with the accuracy of reconstruction limited only by the number of activated primaries actually contacted by the reference neuron. These contacts may be direct or indirect and in any case it is always possible for the supervisory system to utilize more than one reference neuron. Thus it is just as easy to store a complex pattern of activity (e.g. one concomitant to a complicated scene) as a simple one; for all the sensitized neurons which are contacted will be loaded. The memory process is intrinsically super-position-free as long as the supervisory system ensures that no more than one memory is loaded under the control of any reference neuron [since the relation between reference neurons and stored memories is one–one or many–one, cf. postulate (3)]. Note that this relation implies that memories should be accessible through stimulation of specific brain loci (for which there is considerable evidence, cf. Penfield & Perot, 1963), but at the same time is compatible with distributed memory assuming the many–one situation (i.e. redundancy).

The model provides a natural mechanism for memory manipulation. This is rememorization. Suppose that some pattern of neuron firing is called by the reference neuron which previously loaded it rather than resulting directly
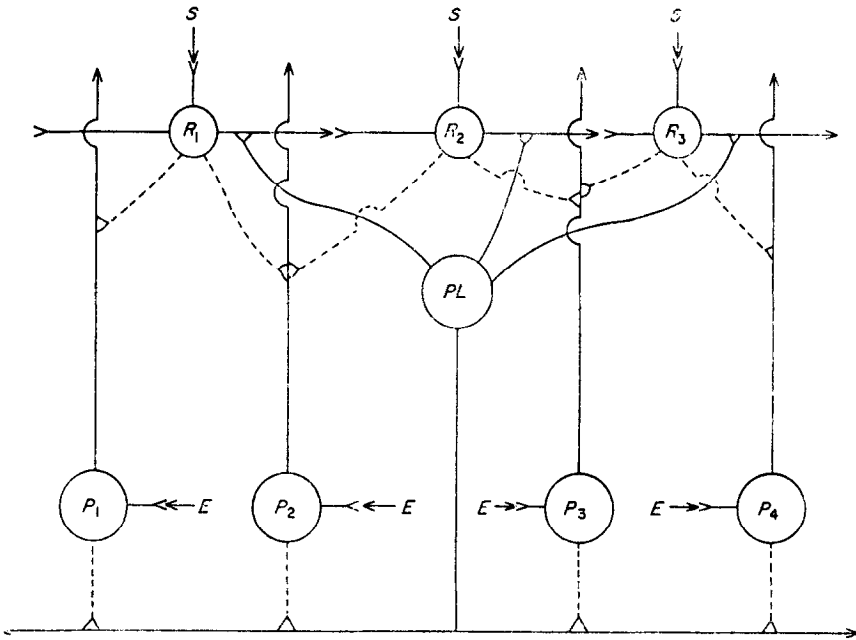
FIG. 2. Schematic of reference neuron circuitry, highly simplified. R_i = reference neuron i; P_i = primary neuron i; PL = party line neuron; E = external, sensitizing input (from preprocessing neurons); S = input from supervisory system. Solid arrows ($\longrightarrow$) are axons, solid tails ($>$—) are unmodifiable dendrites, and broken tails ($>$-------) are modifiable (loadable or callable) dendrites. To trace through the circuitry, suppose that a sequence of external inputs produces firing patterns $P_1 P_2 \bar{P}_3 \bar{P}_4$, $\bar{P}_1 P_2 P_3 P_4$ and $\bar{P}_1 P_2 P_3 \bar{P}_4$ (where P_i represents firing and $\bar{P}_i$ represents non-firing) and that fired neurons in each pattern are loaded by reference neurons R_1, R_2 and R_3, respectively. Activation of these reference neurons for calling reconstructs the patterns of firing at the time when they were active for loading and without superposition. This would not in general be true if any of the reference neurons were used twice, e.g. if R_1 loaded $P_1 P_2 \bar{P}_3 \bar{P}_4$ and $\bar{P}_1 P_2 P_3 P_4$, the reconstructed pattern would be the mixture $P_1 P_2 P_3 P_4$. This type of superposition would be typical if there were no distinction between reference and primary neurons and primaries could only modify or be modified by other primaries. The circuit also illustrates some primary to reference contacts, allowing for content-ordered and associative memory. In general large numbers of both reference and primary neurons are required for efficient content-ordered access.

from external input. This primary pattern can be rememorized under the influence of another reference neuron. Suppose now that two reference neurons are activated for calling, but that some of the output of each is inhibited (which is possible, e.g. if this output reaches the primaries through intermediary neurons). In this case parts of two patterns of firing are reconstructed (e.g. parts of two scenes are recalled). If this new combination is rememorized under the control of a third reference neuron a recomposed (or imagined) scene will be stored in memory.

The concept of rememorization is fundamental. It allows for the construction of sophisticated data structures in the brain. If the supervisory system allows reference neurons to activate one another in series, memories will be recalled in their original time order. By partially suppressed rememorization (as described above) such time ordered structures can be re-time ordered. If the supervisory system allows reference neurons to be activated by primaries, the result is a content ordered memory. Again by rememorization it is possible to load the reference neurons controlling one memory with primaries associated with another memory, thereby establishing an associative structure. Rememorization thus provides a basis for fundamental processes such as classical conditioning and trial and error learning.

A problem with the reference neuron scheme is that large numbers of reference to primary contacts mean a lot of neural wiring. This can be reduced if the system operates on the basis of the party line principle, i.e. if different reference neurons can use the same wires to contact any given primary. In this case the loading signals must be sent in coded form to keep them distinct. The party line principle allows for much increased ramification of reference neuron output, but also requires receptors capable of being separately located. These receptors would presumably be allosteric membrane components (possibly located in different spines) whose conformation records the presence or absence of the loading signals. They must be sensitizable (or have a regulation site which controls the competence to be loaded) and also amenable to some degree of fixation after loading (to prevent unloading by future loading signals). The assumption of sensitization and fixation are necessary even in the absence of party lines; in the presence of party lines it is also necessary to assume individual receptors capable of responding to signal sequences, or rather to the chemical changes (e.g. of transmitter) concomitant to these signals (Conrad, 1974). Recently, Rosen (1976) has shown that such receptors can exhibit the appropriate hysteretic behavior, e.g. by trapping reactants, even assuming only a single steady state under any given conditions.

The model provides a simple explanation for the stability of the memory trace in terms of rememorization-mediated duplication of receptor con-

formations. If reference neurons periodically reload all primaries which respond to their calls and do so before too many receptors lose their conformation and before fixation wears off, the memory will be maintained indefinitely. In principle intracellular potentials could also be used for reloading. If the brain is cooled to extremely low temperatures the receptor conformations (or the openness of the dendrites) will of course be stable without reloading. Thus the model accounts for long-term memory. It also accounts for short-term memory, as memory will automatically be short-term in the absence of rememorization mediated reloadings. To transfer memory from short- to long-term memory it is only necessary to rememorize under the influence of a reference neuron which makes periodic reloads or which contacts dendrites in which fixation is stronger. The fundamental fact about the biochemistry of memory is that inhibition of protein synthesis interferes with long- but not short-term storage (Agronoff, 1970). According to the model this is because long-term reference neurons periodically reload the primary dendrites which they contact, so that the pool of available receptors is always small in these dendrites, from which it follows that interference with protein synthesis will have a much more rapid effect.

The theory provides an interesting interpretation of the histology of the cerebral cortex. This is classically described (Eccles, 1952) as a six-layer structure, with an outer layer of horizontally running axons and inner layers of complexly interconnected neurons. From all the inner layers large neurons (most notably pyramidals) extend apical dendrites into the horizontal layer. These large neurons are presumably the primaries and the horizontal layer the layer of party line telephone wires into which they hook. This is, in fact, the most economical structure for a system operating on the party line principle, although it certainly might be expected to subserve other functions as well. In the cerebellum a similar, but non-party line and much more localized pattern of connectivity is found in the T-like structure of the granule cell through parallel fiber to Purkinje cell circuit (Llinas & Hillman, 1969), where the granule cells can be identified with reference neurons and the Purkinje cells with primaries. This opens the possibility for a memory system or at least a work space for fast calculation.

The essential feature of the reference neuron scheme is the two-tier system of neurons, with triple function of the upper level (reference) neurons. Reference neurons load and call primaries and are also callable by other reference neurons or loadable and callable by primaries. This provides a simple solution to the problem of storing memory without superposition and also provides for a powerful mechanism of memory manipulation and stabilization. This mechanism is rememorization. Variations on the reference neuron scheme are possible and likely from the evolutionary

point of view. However, the psychological concept of rememorization and its neuronal basis in triple function seem to be of potential general significance.

REFERENCES

AGRONOFF, B. W. (1970). In *Molecular Approaches to Learning and Memory* (W. L. Byrne, ed.), pp. 35–39. New York: Academic Press.

CONRAD, M. (1974). In *Physics and Mathematics of the Nervous System* (M. Conrad, W. Güttinger & M. Dal Cin, eds), pp. 108–127. Heidelberg: Springer-Verlag.

ECCLES, J. C. (1952). *The Neurophysiological Basis of Mind.* London: Oxford University Press.

HEBB, D. O. (1949). *The Organization of Behavior.* New York: Wiley.

LLINAS, R. & HILLMAN, D. E. (1969). In *Neurobiology of Cerebellar Evolution and Development* (R. Llinas, ed.), pp. 43–73. Chicago: Am. Med. Ass. Ed. and Res. Fed.

PAVLOV, I. P. (1972). In *Die bedingten Reflex*, pp. 111–121. Munich: Kindler-Verlag.

PENFIELD, W. & PEROT, P. (1963). *Brain* **86**, 595.

ROSEN, R. (1976). *Bull. Math. Biol.* **38**, 87.