Matcher Neural Networks

Peter Z. Revesz

Brown University
Dept. of Computer Science
Providence, Rhode Island 02912

Technical Report No. CS-89-29
April 1989

To appear in Proc. 1st International Joint Conference on Neural Networks, 1989, Washington D.C.

Matcher Neural Networks*

Peter Z. Revesz
Department of Computer Science
Brown University
Providence, Rhode Island 02912

Abstract: This paper presents a powerful new class of neural networks called matcher neural networks. In many ways they are physiologically and psychologically better models for the brain than current connectionist models.

I. Introduction

One intriguing and still largely unresolved problems in psychology is to understand how human beings think. The question is motivated both by human beings' natural curiosity about themselves, and by attempts to find better solutions to problems posed in artificial intelligence and applied in expert systems. Hence all neural models capable of useful function are interesting but those with greatest physiological relevance are most interesting.

Recently a milestone in neural network modeling was reached by the publication of the works of the PDP research group (McClelland and Rumelhart, 1985). The PDP research group has demonstrated algorithms that learn a large class of information, and its models do perform in many situations similar to humans. However, it is puzzling that the group came up with models that in several details contradict the description of the physiological mechanism of the brain as appeared in that very same volume (Crick and Asanuma, 1985). This discrepancy has prompted us to examine the question and as a result to present here a neural network model that does not contradict Crick and Asanuma's description in any major way. At the same time the model exhibits greater psychological validity and lucidity of purpose than currently available models.

II. Assumptions

Neural network models since McCulloch and Pitts' nets attempted to describe the function of the neurons in known logical and mathematical terms:

> Precise specification of these implications by means of recursive functions, and determination of those that can be embodied in the activity of nervous nets, completes the theory. (McCulloch and Pitts, 1943)

and assumed connections that are fundamentally random:

> Genes can only predetermine statistical order, and original chaos must reign over nets that learn, for learning builds new order according to a law of use. (Pitts and McCulloch, 1947)

Soon the model of McCulloch and Pitts became outdated, and the law was interpreted as weight change on the connections by Hebb (1949). Recent years saw many sophisticated extensions of these ideas. For example, in the PDP, each single use of a connection leaves a trace or weight increase which decays at its own exponential rate.

In this paper in contrast to these prevailing ideas or bias, we take a biologically more conservative view, that is, stress efficiency and modularity. In this view individual neurons are efficient modules capable of biologically meaningful function. A group of neurons may form a larger module, which is also biologically meaningful and repeatable. Since modules are repeatable, the organization of the brain need not be as random as previously thought. With this new bias and the aim of discovering the modules we studied the physiology of neurons.

III. Neurons — Physiology

Probably one of the most startling features of neurons is the distribution of their inhibitory synapses. Neurons may have an inhibitory synapse on their spines, on their dendrites, on their soma, on their axon hillock, or even on their axons. This suggests that inhibitory inputs are not all equal. (In contrast, excitatory synapses show more regularity.)

The distribution of inhibitory synapses on pyramidal (spiny) neurons is particularly interesting as described by Crick and Asanuma (1985) who assume that Type I synapses are excitatory and Type II synapses are inhibitory:

> Such cells receive Type I synapses mainly on their spines. Usually, each spine has only a single Type I synapse; a minority of spines has an additional Type II synapse. No spine seems to have only a Type II synapse by itself (Colonnier, 1968).

These facts, in particular the last fact, makes it suspicious that the inhibitory input on a spine inhibits only the excitatory input on the same spine. On a higher level this implies that synapses to neurons are *ordered* and *not*

*This work was supported by NSF grant IRI-8617344.

p	q	out
+	+	▪
+	▪	+
▪	+	-
▪	▪	▪

Figure 1

p	q	out
+	+	▪
+	▪	▪
▪	+	-
▪	▪	▪

Legend: + excitatory, - inhibitory, ▪ no input

Figure 2

permutable as current models assume. That is the major premise of this paper.

Ordering the synapses increases efficiency of the neurons because more information can be conveyed with the same number of inputs. In addition, it enables easy matching of two patterns, where patterns consist of an ordered sequence of +'s and .'s. Matching is an almost unavoidable operation and the linear associator nets essentially do that in a higher level (see Anderson, 1986). Amari (1977), Anderson (1970) and Kohonen (1977, 1984) discuss the merits of linear associators. Here we would like to suggest that matching may be done more directly at the neuron level.

Matching of two patterns can be reduced to comparison. A pattern p contains q, if p has a "+" whenever q does. Therefore building a neuron that generates output only when p contains q is enough. At first assume that the neuron receives an excitatory (enable) signal on its soma. Now enter p_i by an excitatory and q_i by an inhibitory synapse at the same i^{th} spine for each i. Assume the output table shown in Figure 1.

Clearly, when p does not contain q, then there is a q_i which is "+" (present) while p_i is "." (absent). But in this case the soma of the neuron will receive an inhibitory signal from the i^{th} spine, so no output is generated. In any other case, there is at least one excitatory signal to the soma, so an output can be generated.

One intriguing idea is the following. It is possible that p and q are input at different times. Suppose for example that p is entered and some of the spines will become and remain positively charged, enough to neutralize an inhibitory signal, but not strong enough to send any excitatory signal to the soma. This arrangement yields both a means to store information and the more attractive table of Figure 2.

In contrast to pyramidal cells, neurons without spines seem to give only inhibitory outputs, and their action seems triggered by the summated activity of several inputs (Crick and Asanuma, 1985). They are used in the model as simple threshold neurons, that is, they have only excitatory inputs, and they give output only when a certain number or percent of their inputs are firing.

Probably the most surprising feature of our model is that it does not require any weight change on the connections. That weights are necessary is widely believed in spite of the fact that no evidence has been found so far (Crick and Asanuma, 1985). In addition, the working of neurons is often nonlinear. Grossberg (1978) gave probably the best description so far. Matching of course yields a highly nonlinear neuron.

IV. Exemplars and Prototypes — Psychology

The nature of information used and stored is a crucial question in any memory model. In the brain information to and from the cortex flows through the thalamus. The communication between the cortex and the thalamus is exclusively excitatory (Crick and Asanuma, 1985). We suppose then that informations are patterns of neurons that are firing or calm. In an input pattern p, the i^{th} element corresponding to the i^{th} neuron's activity can be noted as "+" for firing and "." for calm. (This is different from McClelland and Rumelhart's inputs which are "+" for exciting and "-" for inhibiting.) For example, when one sees a dog named Fido, a pattern may enter the cortex through a set of neurons. If the i^{th} neuron is firing it may denote something like Fido is short, friendly, red, jumping, has a long tail, or even simpler tidbits about Fido. We assume that between distant parts of the cortex similar patterns of information flow.

Each information pattern sent to the cortex is an exemplar. There are numerous experiments which show human beings' ability to remember and recall exemplars or snap-shot memories (Medin and Schaffer, 1978) and (Brooks et al., 1985). In addition, specific experiences

seem biologically important. Animals can ill afford to expose themselves to the same danger twice. For example, birds may learn fast which berries are good or poisonous.

On the other hand, as Loftus (1977) found, human beings have no preference for exemplars, but when they can they blissfully blend experiences. A funny and illuminating example by Loftus saw people recall a green truck from yellow and distorted blue exemplars. The recall of exemplars and the blending of experiences can be triggered by various attributes in various contexts: human memory is highly content addressable. The nature of blending experiences is not arbitrary. Arguably the pattern that results from blending *is not but is close* to the most typical example of the set of patterns blended (Knapp and Anderson, 1984), and with slight sloppiness it is called a prototype. However, as Knapp and Anderson (1984) showed if the blended pattern is far from any of the exemplars, then people do not respond to it well.

The above implies that in faithful models remembering exemplars is primary and remembering prototypes is secondary. Other authors who share this view include (Medin and Schaffer, 1978) and (McClelland, 1981). Medin and Schaffer (1978) were interested in categorization tasks, which are not topics of this paper. McClelland's (1981) ideas played a prominent part in the PDP models. Posner and Keele (1970) argued strongest against that view based on observations that exemplars fade quicker in memory than prototypes. However, if the exemplars retained form a *small but random* sample of the total, then their findings can be explained within the first view. The combination of the ideas in this and the previous sections leads to a powerful model, which is described in the next three sections.

V. Learning and the Reticular Complex

As Crick (1984) believes and data suggest the reticular complex often directs the information from the thalamus to selected areas of the cortex. This section uses Crick's searchlight hypothesis with modifications to describe a learning algorithm for matcher neural nets. Suppose that a set of n neurons in the thalamus forms successively k patterns to be recorded in the matcher neurons $m_1, m_2, \ldots, m_k$. Let t_i be the i^{th} thalamus neuron. Assume a connection from t_i to the i^{th} spine of each matcher neuron.

To avoid overwriting stored patterns in the matcher neurons the model has to assure that each thalamus neuron is writing only to the same yet unused matcher neuron at any given time. The reticular complex may help exactly this task in the brain. One possible way it may operate follows assuming the organization shown in Figure 3.

Let $r_1, r_2, \ldots, r_k$ and $r'_1, r'_2, \ldots, r'_k$ be the neurons in the reticular complex. Suppose that all r neurons (but not the r' neurons) receive excitatory inputs (some of which may come from the t neurons) at rhythmic intervals. Therefore, normally each r_i will inhibit the connections to m_i. Now suppose that r'_1 receives an excitatory input. Then r'_1 will inhibit r_1, so for the next interval the connections to m_1 will be open and the first pattern will be sent there. However, r'_1 also excites r'_2. Hence in the interval that follows r'_2 will inhibit r_2 and the connections to m_2 will be opened. Note that by that time the connections to m_1 will be closed. Hence when the next pattern comes it will be written to m_2, and so the system may continue always allowing writing to only one matcher neuron.

VI. Recall and Prototype Formation

The above model can be extended, as shown in Figure 4, by adding two more rows of neurons called the f row and the s row. Neuron f_i receives input from t_i without any barrier by the reticular formation. Neuron s_i receives input from t_i one spine at a time. The input sequence to the spines can be controlled by the reticular formation similarly to the previous section.

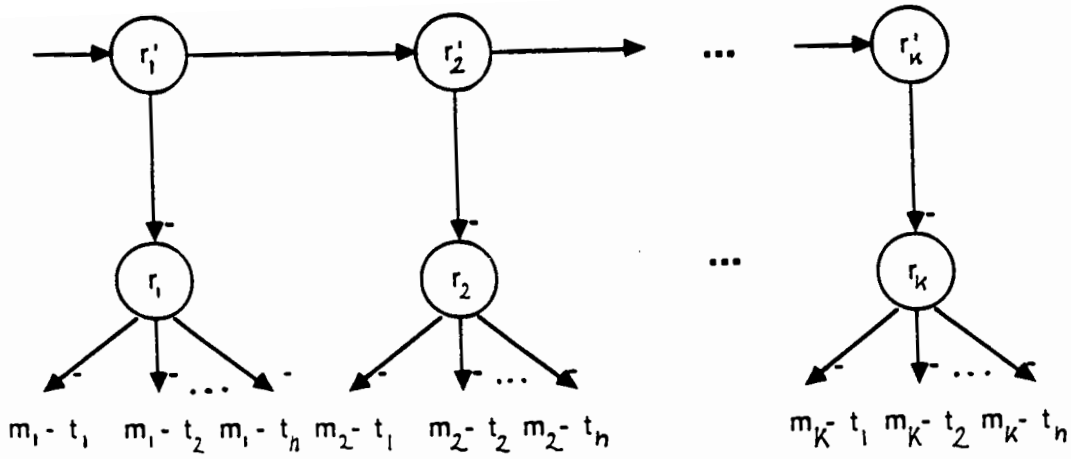
Now suppose that an incomplete pattern like (...+...++....) is input to the area of the cortex which stores information about dogs. The three +'s may mean short, red, and long-tailed, and one of the neurons which is calm may mean friendly. Should that . be completed to a + or left empty? Blending together experiences with short, red and long-tailed dogs tells whether such dogs also tend to be friendly. Therefore, input the pattern by the f row through the inhibitory synapses to the m neurons. If the neurons store information about short, red, and long-tailed dogs, then they will respond sending a pulse to each neuron in the s row.

The inputs to the s neurons are such that s_k receives input from neuron m_i exactly at the spine where the k^{th} element of the i^{th} pattern (t^i_k) is stored. Therefore the spines will generate as many *inhibitory* signals as many patterns are stored that contain the incomplete pattern but do not contain a + at the k^{th} element.

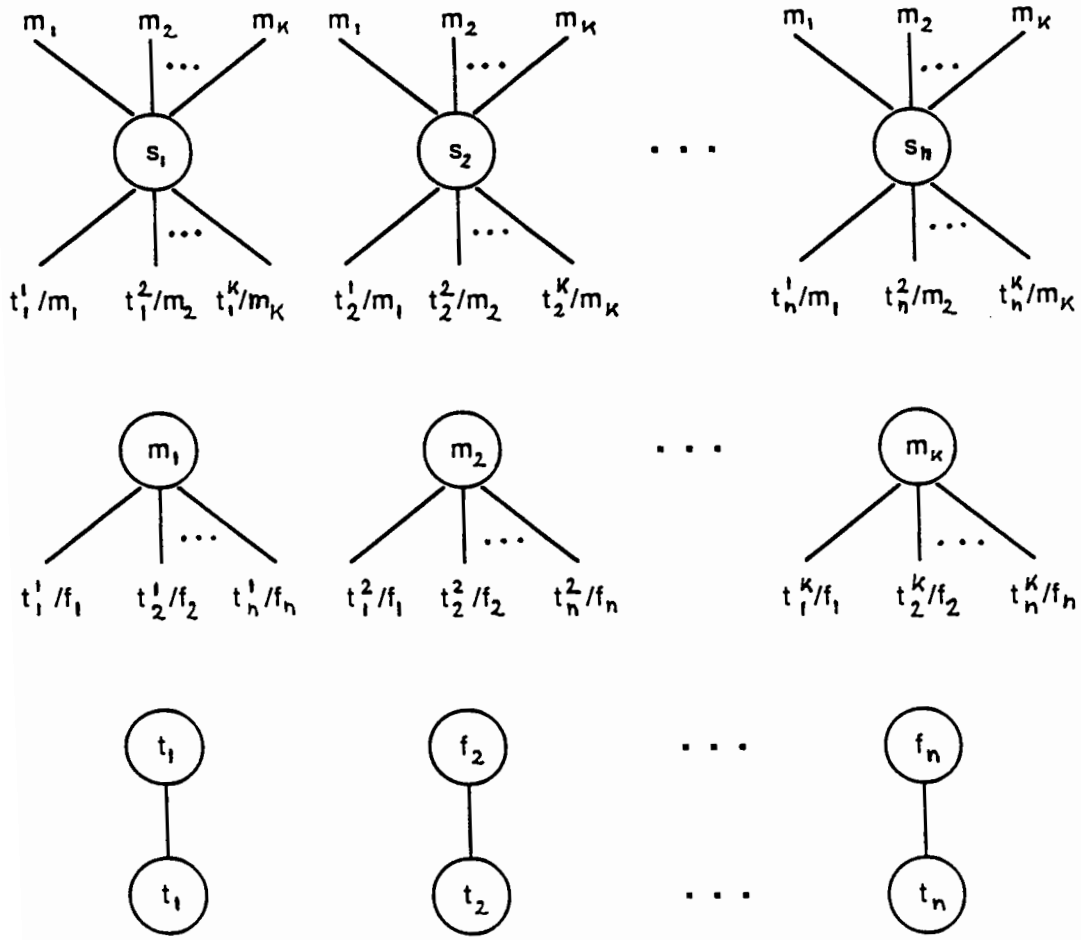
Now suppose that the m neurons also send signals to the soma of the s neurons. Then the soma will also get as many *excitatory* signals as many matches occurred. Suppose that each inhibitory signal is twice as strong as one excitatory signal. Then the neuron s_k will generate output only when p_k is present in more than half of the relevant patterns.

The outputs of the s neurons contain the completed pattern. If friendliness was present in none of the learned relevant patterns, then the s neuron representing friendliness will be calm. Hence foxes can be distinguished from dogs by scant information gained by only a glimpse. Note that the completed pattern is also a prototype for "short, red, long-tailed" dogs.

Enable signals to the m neurons are not shown. They may be the rhythmic diffuse inputs coming from the forebrain. In that case, the forebrain inputs act as timer clicks. During deep sleep, the m neurons would act similarly, but awake they may be acting dissimilarly because they contain different patterns. This is consistent with the well-known findings of (Goldstein et al., 1968), (Noda



A possible organization of the r neurons
Inhibitory inputs are marked by '-'
x - y denotes connection between x and y
FIGURE 3



One module except for the r neurons
t_j^i input from t_j when the i-th pattern is present
x / y denotes excitatory input x and inhibitory input y to same spine
FIGURE 4

and Adey, 1970), and (Hubel and Wiesel, 1962). Goldstein et al. (1968) say that nearby cells in the primary auditory cortex responded in an "individualistic and variegated" manner. Noda and Adey (1970) found a high correlation between nearby cells in the cat's parietal cortex during deep sleep and a weak correlation during REM sleep. On the other hand, the equally well-known findings on distributed representation (Hinton and Anderson, 1981) (Crick and Asanuma, 1985) are also accommodated by the s neurons. In addition, Hirsch and Soinelli's (1971) observations that depriving the cat of visual experience, such as experience of horizontal lines, results in an absence of cells that respond to horizontal lines in the adult can be explained, because it could mean in the model that no matcher neuron stores "horizontal line" in one of its spines. Therefore, when later a visual experience involving "horizontal line" comes, the pattern will not be contained in any of the matcher neurons so they do not respond.

A. Recall with Ambiguous Objects

When seeing an ambiguous picture the sensory perception fluctuates among alternatives. This also may happen when seeing an object only for a flash. Returning to the previous example, this may mean knowing only that the fox is red and short and long-tailed *or* brown and short and long-tailed *or* red and small and long-tailed. How can one find a prototype for an ambiguous object like like that?

This problem is like the previous but with a more complex pattern A, which is a disjunction of ordinary patterns. [1] Finding a prototype is equivalent to finding the probability $p(s_i \mid A)$ for each s_i. It turns out that the same network can be used as before with the following modification of rules: if an m neuron (spiny neuron) fires, then it can not fire again for some brief time. This seems to be biologically true (Crick and Asanuma, 1985).

Then as the small patterns are sent to the cortex in quick succession, each m neuron that contains some of them will send a signal to s_i only once. This means that s_i will have $p(A)$. Moreover, each m neuron will tell by the spine whether it contains the feature that s_i represents. Therefore s_i will also have $p(s_i \cap A)$. This means that s_i will have the information it needs to respond because $p(s_i \mid A) = \frac{p(s_i \cap A)}{p(A)}$. Together the s neurons will give a pattern that best describes the object seen based on previous experience.

VII. Modules, Irregularities, and Variations

The model presented so far is extremely well-structured, yet the brain does not show such clear organization. Any good model then must account for the apparent randomness. A plausible account follows.

[1] The ordinary patterns may be generated by varying the firing rate of neurons as a digitalized means to convey degrees of certainty. More data and theory is needed on this important point (Cooper, 1973).

Let us return to the fox and dog example. The example illustrated how a glimpse of the fox is enough to disambiguate it from a dog, thereby gaining vital information: "not friendly". However, of what use is to perform the disambiguation in reverse, i.e. to tell whether an "unfriendly" animal is red or has a long tail? Clearly, some informations are unlikely to serve as keys while others often do.

Hence, only the key informations from the f row need to pass to the m row, which may well be in the minority as quoted above. Then the inputs which are not paired may also be entered on other parts of the dendrites, or the soma, or be absent. Those present and unmatched seem useless, but such arrangement may still be possible for genetical and evolutionary reasons. Presumably, the system will try to simplify itself and go through a period of apparent randomness. That this is not impossible, consider the following:

> Although neurons without spines may have transient spines on their dendrites early in development, in the mature state they have relatively few, if any, spines. (Crick and Asanuma, 1985)

This suggests that some spiny neurons simplified to nonspiny neurons and not the opposite, that somehow they got spiny. All these suggest a possibility for great variations and randomness. [2]

Still another level of variation is introduced by modules. Mountcastle (1978) already suggested a modular structure for the cortex based on experimental observations. In the model Figure 4 serves as one module. The modules may communicate among themselves by the s and the f rows. A module A for example may receive inputs from module B. While they are still trying to make sense of the same object — seeing different parts of it — module B may discover that it is "purple". So the s neuron representing "purple" may send a pulse to the f neuron of A also representing "purple". Then after a while the s neuron of A representing "angry" may light up and communicate its discovery with other neurons in the brain. [3]

Notice that f neurons may be such that they fire only when several of their inputs are firing. For example, a higher level module's f neuron representing "pet" may wait until inputs are received from all of "animal", "harmless", and "fun". The higher level module would yield associations about "needs to be fed". Notice that features are more complex, more subtle, and require more experience at the higher level modules.

This scheme can account for the connection among our different senses, for very low level modules probably receive most of their inputs from one sense. Moreover, this

[2] Moreover, this fine-tuning to the environment may also explain some of our 'inborn' sensitivity to human faces as opposed to those of polyps.

[3] It is natural to conjecture that chandelier and basket cells (neurons without spines) control the communication among modules similarly in purpose and mechanism to the reticular formation.

can account for the formation of truly abstract concepts, because high level modules always get as input prototypes and do not have to deal with particularities of individual experience.

VIII. Conclusion

We likely overextended our imagination at least in part in explaining the working of the neurons in the brain, and, unfortunately, could not perform experiments. Nevertheless, we hope that in some ways it has pointed in the right direction.

Acknowledgement: I would like to thank James Anderson for helpful comments on an earlier draft of this paper.

References

[1] Amari, S. A. (1977) A mathematical approach to neural systems. In J. Metzler (Ed.) *Systems neuroscience*, New York: Academic Press.

[2] Anderson, J. A. (1970) Two models for memory organization using interactive traces. *Mathematical Biosciences, 8*, 137-160.

[3] Anderson, J. A. (1986) Cognitive capabilities of a parallel system. *Disordered systems and biological organization*, Springer-Verlag.

[4] Brooks, L. R., Jacoby L. L., and Whittlesea, B. W. A. (1985) The influence of specific familiarity on picture identification. To appear.

[5] Colonnier, M. (1968) Synaptic patterns on different cell types in the different laminae of the cat visual cortex: An electron microscope study. *Brain research, 9*, 268-287.

[6] Cooper, L. N. (1973) A possible organization of animal memory and learning. In B. Lundquist and S. Lundquist (Eds.) *Proceedings of the Nobel Symposium on Collective Properties of Physical Systems*. New York: Academic Press.

[7] Crick, F. (1984) The function of the thalamic reticular complex: The searchlight hypothesis, *Proceedings of the National Academy of Sciences, USA, 81*, 4586-4590.

[8] Crick, F., and Asanuma, C. (1985), Certain aspects of the anatomy and physiology of the cerebral cortex. In *Parallel Distributed Processing*, MIT Press.

[9] Goldstein, M. H., Hall L. J., and Butterfield, B. O. (1968), *J. Acoustical Society of America 42*, 444.

[10] Grossberg, S. (1978), A theory of visual coding, memory, and development. In *Formal theories of visual perception*, New York: Wiley.

[11] Hebb, D. O. (1949), *The organization of behavior*, New York: Wiley.

[12] Hinton, G. E., Anderson, J. A. (Eds.) (1981) *Parallel Models of Associative Memory*, Hillsdale, NJ: Erlbaum.

[13] Hubel, D. H. and Wiesel, T. N. (1968), *J. of Physiology 195*, 215.

[14] Knapp, A. G. and Anderson, J. A. (1984), Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 616-637.

[15] Kohonen, T. (1977), *Associative memory: A system theoretical approach*, New York: Springer-Verlag.

[16] Kohonen, T. (1984), *Self-Organization and associative memory*, Berlin: Springer-Verlag.

[17] Loftus, E. F. (1977), Shifting human color memory, *Memory and Cognition, 6*, 696-699.

[18] McCulloch, W. S. and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, *B. Mathematical Biophysics, 5*, 116-133.

[19] McClelland, J. L. (1981) Retrieving general and specific information from stored knowledge of specifics. *Proceedings of the Third Annual Meeting of the Cognitive Science Society*, 170-172.

[20] McClelland, J. L. and Rumelhart, D. E. (Eds.) (1985) *Parallel Distributed Computing, Explorations in the Microstructure of Cognition*, MIT Press.

[21] Medin, D. L. and Schaffer, M. M. (1978) Context theory of classification learning. *Psychological review, 85*, 207-238.

[22] Mountcastle, V. B. (1978) An organizing principle for cerebral function: The unit module and the distributed system. In *The mindful brain*, MIT Press.

[23] Noda, H. and Adey, W. R. (1970) *J. Neurophysiology, 33*, 572.

[24] Pitts, W. and McCulloch, W. S. (1947) How we know universals, the perception of auditory and visual forms, *B. Mathematical Biophysics, 9*, 127-147.

[25] Posner, M. I. and Keele, S. W. (1968) On the genesis of abstract ideas. *Journal of Experimental Psychology, 83*, 304-308.