ON THE STABILITY OF BRAIN-LIKE STRUCTURES

J. S. GRIFFITH
From the Chemistry Laboratory, Cambridge, England

ABSTRACT Previous authors have argued that the maintenance of the highly connected aggregates of nerve cells in the central nervous system in a stable state of intermediate activity presents something of a paradox. In the present paper it is shown that this is not so and that either a relatively large-scale structure of the aggregate or the presence of inhibitory connections makes a stable intermediate activity possible. It is suggested that large-scale structure can usefully be discussed from an information-theoretic viewpoint and that it is also related to the ergodic problem of classical mechanics.

1. INTRODUCTION

It was shown by Beurle (1956) that a mass of units capable of emitting regenerative pulses would, at least in some circumstances, have an inherently unstable activity. That is, his mass had the property that if it was started in an intermediate state of activity it would shortly become either completely quiescent or completely active. If a real or artificial brain were composed of such units, it would therefore rapidly pass to one of two states which could conveniently be termed deep coma or epilepsy. Hence if such a brain were observed to have usually an apparently stable intermediate activity one would be forced to suppose the existence of some powerful regulating mechanism.

Recently Ashby, Von Foerster, and Walker (1962) have discussed this matter further. They used a slightly different model (similar to that of Rashevsky, 1945) having a related underlying structure and obtained the same property of instability. They went on to suggest that natural brains, which normally operate in an intermediate range of activity, offer something of a paradox. In the present paper two modifications of their model are considered. In both of these the alteration is, I believe, in a direction which is physiologically plausible and in both it is shown that an intermediate and stable activity is possible for the mass of units.

Before continuing, let me emphasize that there are at least two reasons for being interested in this problem. One is for the light it may cast upon the mode of operation of natural brains. The other is that we need to understand these things when trying to construct artificial systems having behaviour similar to or as complex as

natural brains. This latter application is, potentially at least, quite as interesting as the former.

2. THE NATURE OF THE MODEL

We now discuss the model used by Ashby *et al.* (1962). It is supposed that there is a mass, or network, of identical units each having n inputs from other units. The approximation is made of quantizing time by splitting it up into consecutive intervals of length Δt. A unit fires in one of these intervals if and only if at least θ of the units attached to its inputs have fired in the preceding interval. θ is the threshold of a unit. Clearly this quantization would be strictly true if, as in a present-day digital computer, there were a train of synchronizing clock pulses passing into the mass, but it is presumably incorrect for animal brains. Inasmuch as one believes, then, that the present model is relevant to animal brains one must watch out for errors arising from this approximation. We return to this point in section 7.

Ashby *et al.* further supposed that the activity of a mass in a time interval may be adequately represented by a single number p satisfying $0 \leq p \leq 1$. p is the probability that a unit fires in the interval. We shall write $P(p)$ for the corresponding quantity for the net interval of time It is calculated by considering one unit with its n inputs. The probability that the unit attached to any chosen input has fired in the previous time interval is taken as p. Furthermore it is assumed that the probabilities for the n inputs are independent. Hence the probabilities for various numbers of inputs to have been active are given by the binomial distribution. $P(p)$ is the probability that at least θ of the inputs were active, *i.e.*:

$$P(p) = \sum_{N \geq \theta} \binom{n}{N} p^N (1 - p)^{n-N} \tag{1}$$

The instability follows from this formula and apart from one point of unstable equilibrium, p tends to 0 or 1.

The instability can also profitably be considered from an information-theoretic viewpoint. The block of material has two stable states, therefore it can store 1 bit of information only. Thus, in one way, its useful complexity of structure is no greater than that of a single bistable circuit. Alternatively, if it is thought of as an information-processor, any initial state gets processed into one of two states—or in other words only one bit of information can be extracted from an arbitrary initial state. In my opinion, this inability to retain information is the most significant fact about the instability.

We now pass on to consider two modifications to the treatment of Ashby *et al.* The first is in the next section and involves examining and rejecting the assumption that a fairly richly connected network can necessarily be adequately described in terms of a single quantity p. Then in the remaining sections the effect of introducing inhibitory as well as excitatory connections between units is discussed.

3. ERGODIC AND NON-ERGODIC NETWORKS

We lead into our first modification by considering a special network whose properties are easily calculable without making any approximations. Suppose the network contains just n units, each unit being one of the n inputs for each of the other units and for itself. We shall call this a complete network. Evidently, if at one interval of time m of the units fire, then at the next interval no units or n units fire depending on whether $m < \theta$ or m $\geq \theta$ respectively. Thus a complete network jumps immediately into one of the two extreme activities. Note that if we tried to apply formula (1) to this situation we would put $p = m/n$ and predict the subsequent number of active units to be $nP(p)$ which is not generally equal to 0 or n. The reason for this discrepancy is that, although a unit selected at random has probability p of firing, the probabilities for the n units are obviously not independent.

More significantly for the construction of a brain, one could make a transmission line out of this material. This would be done by placing sets of n units in a linear chain as illustrated for $n = 3$ in Fig. 1. Here each of the n units at one stage serves

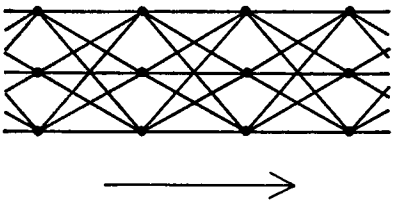

FIGURE 1 Part of a complete transmission line for $n = 3$.

as an input to each of the n units at the stage next to the right. We shall call this a complete transmission line. It has the property that it will accept any of $n!$ possible inputs and transmit reliably one bit of information about the input, namely, whether $m < \theta$ or $m \geq \theta$.

At this point one may well ask if there is any advantage in having so many units at each stage. Would it not be sufficient to contract the line to one unit after the first stage as in Fig. 2? After all, one only needs one unit per stage to conduct one bit of

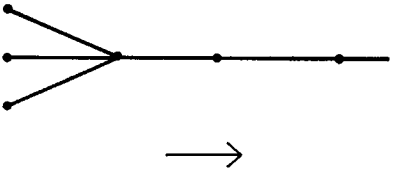

FIGURE 2 Convergent transmission line. $n = \theta = 1$ after the first stage.

information. If the individual units are reliable, the answer to this is yes. However if they are not there is an advantage in having several units per stage. Clearly, the

larger n the more errors in single units can be tolerated. Also, numbers of units may fail completely without the line ceasing to transmit correctly. Thus by sacrificing complexity of task, a complete transmission line gains great reliability. It is well known that animal brains do not appear to have their operation significantly altered by removal of quite large pieces of material from many parts of the cortex; obviously some form of reliability by redundancy must obtain there, although there is no evidence to suggest whether it is achieved by the kind of mechanism discussed here (the present mechanism is similar to the "multiple line trick" of Von Neumann, 1956).

Next suppose we place x complete transmission lines side by side but with no connections between them. Clearly this gives us a composite line capable of transmitting reliably x bits of information. This already refutes the conclusion obtained by arguing in terms of the average activity p, namely, that there are necessarily only the two stable extreme activities. Our composite line has $x + 1$ different stable values for p ranging from $p = 0$ by steps of x^{-1} up to $p = 1$.

There is a considerable analogy between the present situation and the ergodic problem of classical mechanics. In the latter, one is concerned with the question: when is the long-term time average of a property of a dynamical system equal to the average of that property over the phase space of the system? Now we are concerned with whether we get the right long term behaviour of our network by considering only the temporal development of the average quantity p. In classical mechanics the ergodic property is true almost everywhere in phase space if the phase space is metrically indecomposable (Khinchin, 1949). For our network we have shown a weak converse of this, namely, that it is definitely incorrect to argue only in terms of p if the network can be decomposed into two or more disconnected parts.[1] Consequently we shall term ergodic a scheme which works in terms of the quantity p as described in section 2.

It is certainly premature to have any opinion as to whether, when an animal brain extracts x bits of information through a sensory input apparently capable of carrying much more, it could be doing so by a mechanism involving x complete transmission lines. However, it is reasonable to enquire what kind of experiment might give some evidence about this. The most obvious would seem to be the histological one of actually observing a separation between the x lines. Yet while this might give positive evidence of non-ergodicity it would be much more difficult to show the converse. This is so for at least two reasons. The separation of the units into lines is functional and not necessarily geometrical. That is, the lines could be intertwined in a quite haphazard way. Also, although we assumed our x lines to be completely disconnected, they could have considerable cross-linking without destroying their ability to carry x bits. This follows from the great stability, in a complete network, of

[1] Compare Elsasser, 1962. Each bit of information may be regarded as a temporary constant of the motion, in his terminology.

the quiescent and fully active states. It is presumably to the advantage of a brain to be able to alter the details of the distribution of the xn units in x subsets, either by excitatory and inhibitory control pulses from elsewhere in the brain or perhaps by growth or decay of connections. Hence one might well expect to find such cross-connections.

The more hopeful method would seem to be through the mathematical analysis of microelectrode measurements (see for example, Gerstein and Kiang, 1960, Rodieck, Kiang and Gerstein, 1962). A complete transmission line must be either completely quiescent or completely active, at any stage beyond the first.[2] However when we include inhibition we shall see that this property disappears. It is also possible to measure the joint activity of pairs of cells by this method and to analyse it for the presence of correlation. Some experimental results are available for the visual cortex of cat, showing correlation in many but not all cases, but are too few yet to enable any useful deduction to be drawn about the connectivity (Griffith and Horn, 1963). Ultimately, however, this type of measurement should tell us a lot about the connectivity and hence ergodicity of natural brains. In particular, the activities of a pair of cells from the total of xn at a stage of the composite transmission line would be correlated if they belonged to the same complete component and uncorrelated otherwise. If the pair were selected at random, the probability of their being correlated would be $(n-1)/(xn-1)$.

4. THE NET WITH INHIBITION

We now generalize the ergodic approximation to include inhibitory inputs. Specifically we suppose each cell to have n_1 excitatory and n_2 inhibitory inputs. The threshold is still written θ and we suppose the cell fires if the numbers N_1 of excitatory and N_2 of inhibitory pulses in the preceding instant of time satisfy the relation

$$N_1 - \phi N_2 \geq \theta \tag{2}$$

Here ϕ is a positive constant, which need not be integral. The linear inequality (2) is clearly not the most general plausible threshold relation possible but it is sufficiently general for our purpose. In particular if we choose $\phi > n_1 - \theta$ we have the case where even a single inhibitory pulse prevents the cell from firing, no matter how many excitatory pulses occur simultaneously. Apart from allowing ϕ to be non-integral, the present net is a neural net with relative inhibition as considered by McCullough and Pitts (1943).

We still hopefully make the ergodic assumption that the over-all state of the net at a given time may be adequately represented by a single average probability p of firing of a randomly selected cell in the unit of time Δt. Then the average probability for the next unit of time is

[2] Hence a microelectrode should reveal zero or maximum activity.

$$P(p) = \sum_{N_1 - \phi N_2 \geq \theta} \binom{n_1}{N_1}\binom{n_2}{N_2} p^{N_1+N_2} q^{n_1+n_2-N_1-N_2} \tag{3}$$

where $q = 1 - p$ and the sum is over all N_1 and N_2 satisfying the threshold relation. We also define $R(p)$ by

$$\begin{aligned} R(p) &= p^{-1}P(p) \qquad p \neq 0 \\ R(0) &= \lim_{p \to 0} R(p) \end{aligned} \tag{4}$$

Evidently, for $0 < p < 1$, the activity of the net at the later instant of time is greater or less than that at the earlier depending on whether $R > 1$ or $R < 1$ respectively. When $R = 1$, the corresponding value of p gives a stationary, though not necessarily stable, activity for the net (these stationary points have similar significance in this model to the fix-points of Von Neumann, 1956, p. 97, and the critical points occurring in the author's field theory of neural nets (Griffith, 1963)).

The behaviour of the net now depends directly on the form of R as a function of p and indirectly on the two structural parameters n_1, n_2 and the two threshold parameters θ and ϕ. In the next section we show that for an extensive class of parameters with large $n = n_1 + n_2$, the net still has only two stable activities, namely, with $p = 0$ or $p = 1$. Then in the following section we show that there is another class for which there is a further stable activity with an intermediate value of p.

5. THE CASE OF n_1 AND n_2 LARGE

When we have a large number of both excitatory and inhibitory inputs, the most obviously interesting situation occurs with a large threshold comparable with both n_1 and n_2. Therefore we write $n_1 = \mu_1 n$, $n_2 = \mu_2 n$, $\theta = tn$, where μ_1, μ_2, t are numbers between 0 and 1. ϕ is also a small number, not necessarily an integer, and n is large. The advantage of this formulation is that it enables us to use de Moivre's theorem, which is a prototype of the central limit theorem, to obtain a simple approximation to P of equation (3) which is asymptotically accurate as $n \to \infty$.

To apply the theorem we write (Cramér, 1955)

$$N_1 = n_1 p + \lambda_1 \sqrt{npq}$$

and similarly for N_2. Then the probability that λ_1 lies between x_1 and y_1 is asymptotically

$$\frac{1}{\sqrt{2\pi}} \int_{x_1}^{y_1} e^{-(1/2)t^2}\, dt$$

whence

$$P(p) \doteqdot \frac{1}{2\pi} \iint_C e^{-(1/2)(\lambda_1^2 + \lambda_2^2)}\, d\lambda_1\, d\lambda_2 \tag{5}$$

where the integration is over the region C in the λ_1, λ_2 plane satisfying

$$\lambda_1\sqrt{\mu_1} - \phi\lambda_2\sqrt{\mu_2} \geq \left(\frac{n}{pq}\right)^{1/2}(t - \mu_1 p + \phi\mu_2 p) \tag{6}$$

As $n \to \infty$ the right-hand side of equation (6) tends to $+\infty$, 0 or $-\infty$ as $t - \mu_1 p + \phi\mu_2 p$ is greater than, equal to, or less than 0, respectively. In each case the integral (5) is easily evaluated and we find

$$\left.\begin{aligned} P(p) &\doteqdot 0 \quad \text{when} \quad p < \frac{t}{\mu_1 - \phi\mu_2}, \\ P(p) &\doteqdot \tfrac{1}{2} \quad \text{when} \quad p = \frac{t}{\mu_1 - \phi\mu_2}, \\ P(p) &\doteqdot 1 \quad \text{when} \quad p > \frac{t}{\mu_1 - \phi\mu_2}. \end{aligned}\right\} \tag{7}$$

Equations (7) assume $\mu_1 > \phi\mu_2$. If $\mu_1 < \phi\mu_2$, $P(p) \doteqdot 0$ for all p.

Clearly $R(p)$ is practically zero up to near the point $p = t(\mu_1 - \phi\mu_2)^{-1}$ where it rapidly changes to p^{-1}, which is greater than one, up to $p = 1$. Thus there is one intermediate value of p with $R(p) = 1$, but it is unstable. As in the case considered by Beurle (1956) and Ashby *et al.* (1962) the only stable activities are for $p = 0$ or 1.

6. THE CASE $n_2 = 1$

Another class of possibilities is given by the assumption that n_1 is large; n_2, small; and that the inhibitory links are very strong in order to make up for their small number. We consider specifically $n_1 = n$; $n_2 = 1$; θ, a small integer; ϕ, large. Again we let $n \to \infty$ with either $\phi = \kappa n$ or $\phi = n - \kappa$. These somewhat special assumptions are made so that we can do the necessary mathematics easily and transparently and will be shown to lead to the conclusion that we do, in general, get a stable activity for some p not equal to 0 or 1.

We first demonstrate the possibility of this stable intermediate activity by restricting the case $\phi = n - \kappa$ even further by requiring $\kappa = \theta$. Then equation (3) becomes

$$P(p) = 1 - p + \sum_{i=0}^{\theta}\binom{n}{i}[-p^i(1-p)^{n-i+1} + p^{n-i+1}(1-p)^i] \tag{8}$$

and P satisfies

$$P(p) + P(1 - p) = 1. \tag{9}$$

Hence $P(\tfrac{1}{2}) = \tfrac{1}{2}$, $R(\tfrac{1}{2}) = 1$ and so the point $p = \tfrac{1}{2}$ is stationary for any n. It will be natural to call it stable if for nearby values of p, $R(p) < 1$ for $p > \tfrac{1}{2}$ and $R(p) > 1$ for $p < \tfrac{1}{2}$. This is so if $(dR/dp) < 0$ at $p = \tfrac{1}{2}$. It follows from (8) that

$$\left(\frac{dR}{dp}\right)_{1/2} = -4 + \left(\frac{1}{2}\right)^n \sum_{i=0}^{\theta}(n - 2i + 1)\binom{n}{i}. \tag{10}$$

The right-hand side of equation (10) clearly tends to -4 as $n \to \infty$. For fixed θ, therefore dR/dp is negative for all sufficiently large n but not necessarily for small n. We have now established that there is generally a stable point at $p = \frac{1}{2}$. Furthermore, for any fixed $p \neq 0, 1$,

$$\lim_{n\to\infty} P(p) = 1 - p, \qquad \lim_{n\to\infty} R(p) = p^{-1} - 1.$$

We now consider $\phi = \kappa n$, with $0 < \kappa < 1$. Then

$$R(p) = (p^{-1} - 1) \sum_{N_1 \geq \theta} \binom{n}{N_1} p^{N_1}(1-p)^{n-N_1} + \sum_{N_1 \geq \theta + \kappa n} \binom{n}{N_1} p^{N_1}(1-p)^{n-N_1} \qquad (11)$$

Clearly as $p \to 0$, $R(p) = 0(p^{\theta-1})$. So $R(0) = 0$ providing $\theta > 1$. Similarly $R(1) = 1$ and $(dR/dp)_1 = -1$ providing that $n \geq (\theta + 1)/(1 - \kappa)$. So both $p = 0$ and $p = 1$ are stable.

For intermediate values of p we again use de Moivre's theorem and find

$$R(p) \doteqdot \frac{1-p}{p\sqrt{2\pi}} \int_{C_1} e^{-(1/2)x^2}\, dx + \frac{1}{\sqrt{2\pi}} \int_{C_2} e^{-(1/2)x^2}\, dx \qquad (12)$$

where C_1 is the range $x\sqrt{npq} \geq \theta - np$ and C_2 is $x\sqrt{npq} \geq \theta + n(\kappa - p)$. It easily follows that as $n \to \infty$, $R \to p^{-1} - 1$ for $p < \kappa$, $R \to \frac{1}{2}(p^{-1} - 1)$ for $p = \kappa$ and $R \to p^{-1}$ for $p > \kappa$. Hence $p = \frac{1}{2}$ is again a stable point, providing $\kappa > \frac{1}{2}$. There are also two unstable stationary activities near $p = 0$ and $p = \kappa$.

As an illustration, R is plotted in Fig. 3 as a function of p for two particular cases. In Fig. 3*a*, $n_1 = 4$, $n_2 = 0$, $\theta = 2$, $\phi = 1$ and

$$R(p) = 6p - 8p^2 + 3p^3$$

We see the two stable and one unstable stationary points, as discussed by Ashby *et al.* With an inhibitory link, however, a stable intermediate activity is possible even for small n_1. Figure 3*b*, with $n_1 = 5$, $n_2 = 1$, $\theta = 2$, $\phi = 2$ and

$$R(p) = 10p - 30p^2 + 35p^3 - 14p^4$$

illustrates this (*see top of next page*).

7. LONG-TERM BEHAVIOUR OF NETWORK

In the last section we found that in a number of cases $p = \frac{1}{2}$ was a stable point surrounded by a considerable range of p for which $P(p) \doteqdot 1 - p$. We now assume $P(p) = 1 - p$ and ask what will be the long-term behaviour of the network. Starting with activity p in the first instant of time, the activity is $1 - p$ in the second instant and $1 - (1 - p) = p$ in the third. So we have the surprising result that over a wide range of initial values p, the network oscillates between the activities p and $1 - p$. As $P(p)$ is not accurately $1 - p$, the range of the oscillation will slowly alter.

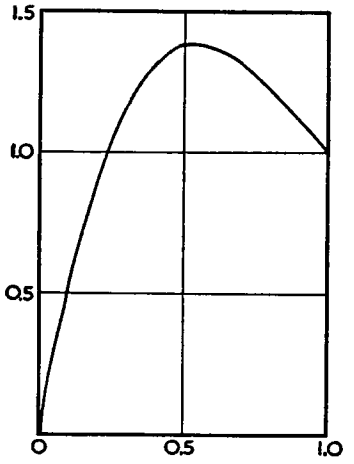

FIGURE 3a

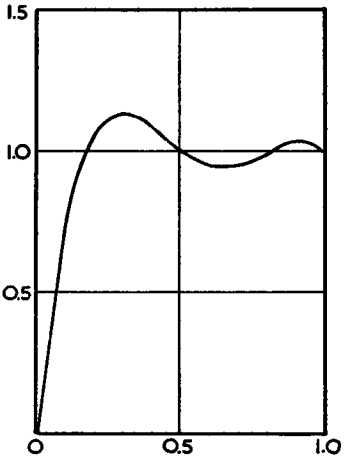

FIGURE 3b

FIGURE 3 R plotted as a function of p. (a) $n_1 = 4$, $n_2 = 0$, $\theta = 2$, $\phi = 1$; (b) $n_1 = 5$, $n_2 = 1$, $\theta = \phi = 2$.

We have now reached a conclusion we decided to beware of in section 2. Our time intervals are quantized, somewhat arbitrarily, and we finish up with an oscillation having twice the period of the time interval. Our conclusions would apply to a system having synchronizing clock pulses; we have shown that the mass tends to oscillate with a period twice that of the clock. It is hoped to investigate in detail what may happen in a system without clock pulses in a later paper but we shall include one such system here.

Suppose the activity p at time t is given not by $P(t - \tau)$, where τ is our time interval, but by an integral of $P(t)$ over all times previous to t. Specifically let us write

$$p(t) = \int_{-\infty}^{t} P(u)i(u - t)\, du \tag{13}$$

The function i was discussed elsewhere (Griffith, 1963) and is to be regarded as a sort of collection function. It satisfies $i(x) = 0$ for $x > 0$ and

$$\int_{-\infty}^{\infty} i(x)\, dx = 1$$

Our synchronized situation corresponds to $i(x) = \delta(x + \tau)$ where δ is Dirac's δ function. We shall now take the exponential expression $i(x) = \lambda e^{\lambda x}$ for $x \leq 0$. Evidently $i'(x) = \lambda i(x)$.

Putting $P(u) = 1 - p(u)$ in equation (13) and using this $i(x)$ we find

$$p'(t) = P(t)i(0) - \int_{-\infty}^{t} P(u)i'(u - t)\, du$$

$$= \lambda(1 - p(t)) - \lambda p(t) = \lambda(1 - 2p(t))$$

whence

$$p(t) = \tfrac{1}{2} + (p(t_*) - \tfrac{1}{2})e^{2\lambda(t_*-t)} \tag{14}$$

Hence $\lim_{t\to\infty} p(t) = \frac{1}{2}$ and the oscillation is averaged out.[3]

We have now completed our demonstration of the possibility of a stable intermediate activity in a richly connected mass of regenerative units with threshold. This means that such masses are potentially useful for the construction of artificial brain-like systems. On the other hand, although we have not found any properties which are obviously inconsistent with the behaviour of natural brains, we should not regard the present investigation as affording any significant *a posteriori* evidence that neurones actually satisfy firing rules at all analogous to those assumed here. It is very likely that many other rules would give very similar large-scale behaviour to suitably constructed large networks. We have, however, refuted the suggestion that the usually moderate activity of natural brains necessarily presents a paradox.

Received for publication, February 26, 1963.

REFERENCES

1. ASHBY, W. R., VON FOERSTER, H., and WALKER, C. C., *Nature*, 1962, **196**, 561.
2. BEURLE, R. L., *Phil. Tr. Roy. Soc. London, Series B*, 1956, **240**, 55.
3. CRAMÉR, H., The Elements of Probability Theory, New York, John Wiley & Sons, Inc., 1955.
4. ELSASSER, W. M., *J. Theoret. Biol.*, 1962, **2**, 164.
5. GERSTEIN, G. L., and KIANG, N. Y.-S., *Biophysic. J.*, 1960, **1**, 15.
6 GRIFFITH, J. S., *Bull. Math. Biophysics*, 1963, **25**, 111.
7. GRIFFITH, J. S., and HORN, G., 1963, data to be published.
8. KHINCHIN, A. I., Mathematical Foundations of Statistical Mechanics, New York, Dover Publications, Inc., 1949.
9. MCCULLOUGH, W. S., and PITTS, W., *Bull. Math. Biophysics*, 1943, **5**, 115.
10. RASHEVSKY, N., *Bull. Math. Biophysics*, 1945, **7**, 223.
11. RODIECK, R. W., KIANG, N. Y.-S., and GERSTEIN, G. L., *Biophysic. J.*, 1962, **2**, 351.
12. VON NEUMANN, J., *in* Automata Studies, (C. E. Shannon, and J. McCarthy, editors) Princeton, Princeton University Press, 1956, 43.

[3] The block of cells is thus a crude form of restoring organ in the sense of Von Neumann, 1956.