U. S. AIR FORCE

PROJECT RAND

RESEARCH MEMORANDUM

REPRESENTATION OF EVENTS IN NERVE NETS AND FINITE AUTOMATA

S. C. Kleene

RM-704

15 December 1951

Assigned to _____

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

TABLE OF CONTENTS

Page No.

INTRODUCTION. 1

1. Stimulus and Response. 1

2. Nerve Nets and Behavior. 2

PART I — NERVE NETS. 6

3. McCulloch-Pitts Nerve Nets 6

4. The Input to a Nerve Net 8

5. Definite Events. 10

5.1 "Definite events" defined. 10

5.2 Definite positive events 11

5.3 Simpler nerve nets 14

5.4 Definite events in general 19

5.5 Representation of events in general. . . . 24

5.6 Nerve nets without circles 26

6. Indefinite Events—Preliminaries 28

6.1 Some examples. 28

6.2 Initiation 30

6.3 Definite events reconsidered 36

6.4 Why consider indefinite events?. 43

7. Regular Events 46

7.1 "Regular events" defined 46

7.2 An algebraic transformation. 52

7.3 Identity and equivalence 55

7.4 Representability of regular events 62

7.5 Discussion of the proof and further problems 68

7.6 Conjunction and negation. 73

PART II — FINITE AUTOMATA 75

8. The Concept of a Finite Automaton 75

8.1 Cells 75

8.2 State 77

9. Regularity of Representable Events. 80

APPENDIX 1: DEFINITENESS OF EVENTS REPRESENTABLE IN A FINITE AUTOMATON WITH AN INFINITE PAST. . . . 87

APPENDIX 2: PRIMITIVE RECURSIVENESS OF REGULAR EVENTS . . 90

APPENDIX 3: AN EXAMPLE OF AN EVENT WHICH IS NOT REPRESENTABLE, THOUGH IT IS PRIMITIVE RECURSIVE. . 95

BIBLIOGRAPHY 98

Summary: To what kinds of events can a McCulloch-Pitts nerve net respond by firing a certain neuron? More generally, to what kinds of events can any finite automaton respond by assuming one of certain states? This memorandum is devoted to an elementary exposition of the problems and of results obtained on it during investigations in August 1951.

REPRESENTATION OF EVENTS IN NERVE NETS AND FINITE AUTOMATA

S. C. Kleene

INTRODUCTION:

1. Stimulus and Response: An organism or robot receives certain stimuli (via its sensory receptor organs) and performs certain actions (via its effector organs). To say that certain actions are a response to certain stimuli means, in the simplest case, that the actions are performed when those stimuli occur and not when they do not occur.

Since both the stimuli and the actions may be very complicated, the relationship between the two is very complicated.

In order to simplify our analysis, we may leave out of account the complexities of the response. To do this, we reason that any kind of stimulation, or briefly, any event which affects action, in the sense that according as the event occurs or does not, under some set of other circumstances held fixed, a different action ensues, must have a representation in the state of the organism or machine, after the event has occurred and prior

to the ensuing action (which action may depend on the occurrence of many other events).

We then ask what kinds of events are capable of being represented in the state of the organism or machine.

We shall see later (Section 5.5) that there is no loss of generality in considering the representation, in the case of nerve nets, to have the simple form of the firing (or sometimes the non-firing instead) at a certain time of a certain neuron.

For explaining response as due to stimulus, it would then remain to assemble the complicated molar response out of these molecular representations of molar stimuli.

In this remaining problem, it could make a great difference what events are selected for molecular representation, as the abstract from experience which is to form the basis of action.

However, we shall not enter into this here, except as it reflects on the problem of representing events; nor shall we enter into the analogies between the analysis just described and the psychological phenomena in which raw sense data lead through percepts and concepts to overt behavior.

2. Nerve Nets and Behavior: McCulloch and Pitts (1943) in their fundamental paper on the logical analysis of nervous activity formulated certain assumptions which we shall recapitulate below (Section 3).

These assumptions are an abstraction from the data which neurophysiology provides. The abstraction gives a model, in

terms of which it becomes an exact mathematical problem to see what kinds of behavior the model can explain. The question is left open how closely the model describes the activity of actual nerve nets; and some modifications in the assumptions lead to similar models. Neurophysiology does not currently say which of these models is most nearly correct—it is not plausible that any one of them fits exactly. It is noteworthy, however, that one of McCulloch and Pitts' results is that these several other models are capable of producing only the same behavior as the first one.

Until neuro-physiology tells us more about the actual process, it is instructive to see what behavior the model admits. Our results are to the effect that "it could be this way, and quite possibly the real process is significantly similar to this." Furthermore, such studies have applications in robotology, when we wish to describe on paper (or build in the metal, using elements which behave like McCulloch-Pitts neurons) a robot to behave in a pre-assigned manner.

This study can be pursued on two levels, a strictly practical one and a theoretical one. On the former, we are concerned with constructing particular nerve nets to give particular described behavior; in the latter, we develop general methods for constructing nets to give behavior, and investigate the limitations within which this is possible.

This memorandum deals with studies on the second level, but actually the two are not clearly separated. The general methods

may be practical or suggest methods which are, and the investigations of the limitations may contribute better understanding of the problems which are faced on the practical level.

McCulloch and Pitts give such a theoretical investigation, consisting of a theory for nerve nets without "circles" (Part II of their paper) and a theory for arbitrary nerve nets (Part III). The present memorandum is partly an exposition of the McCulloch-Pitts results; but we found the part of their paper which treats of arbitrary nets obscure; so we have proceeded independently here.

Under the McCulloch-Pitts assumption of the all-or-nothing character of a neuron's firing (which is close to the biological reality) and their assumption which quantizes time so that all neurons have their moments of possible firing in phase, a nerve net has the character of a digital automaton. Here we are using "digital" in contrast to "analog," in the sense familiar in connection with computing machines.

It seems quite clear that many physical processes of control are partly analog in character. For example, the respiratory cycle of activity can be controlled consciously(by nervous means, which are digital); but most of the time it is regulated by a nervous response in the respiratory center of the brain to the carbon-dioxide level in the blood (an analog quantity).

Just as in mathematics continuous processes can be approximated by discrete ones, it is plausible that any analog elements

in bodily control could be approximated in their effect by digital ones. Nevertheless, the analog or partly analog controls may remain the simplest and most efficient.

One of the results of systematic theoretical investigations of the potentialities of digital control might be to demonstrate that other principles, e.g., analog mechanisms or the introduction of random inputs, may be necessary to produce, or to produce economically, certain kinds of behavior.

Another tacit assumption of the present mathematical theory is that there are no errors in the functioning of neurons; i.e., a given neuron fires at a given moment, if and only if it should under the McCulloch-Pitts rules. Of course, this is unrealistic, either for living neurons or for the equivalent units of a mechanical automaton. It seems natural, however, to build a theory of what happens assuming no malfunctioning. In this theory, we represent the occurrence of an event by the firing of a single neuron. Biologically, it is implausible that important information should ever be represented in an organism in this way. But by duplications of nets (many processes being carried out in parallel circuits), one could expect then to secure the same results with small probability of failure in nets constructed of fallible neurons.

Returning to the formulation of the problem as given in Sect. 1, we shall now in Part I show that all events of a certain class can be represented by the firing (or in some cases, the non-firing) of a certain neuron. The discussion of the

converse is left mainly to Part II, where we generalize to representability in any finite digital automaton.

PART I — NERVE NETS:

3. McCulloch-Pitts Nerve Nets: Under the assumptions of McCulloch and Pitts (1943), a nerve cell or neuron consists of a soma, whence nerve fibers (axons) lead to one or more endbulbs.

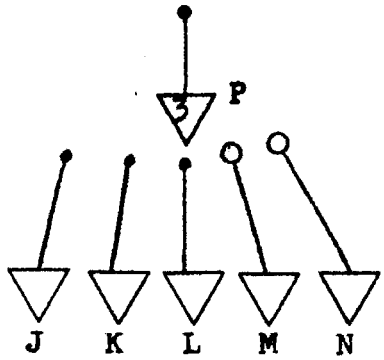
A nerve net is an arrangement of a finite number of neurons, in which each endbulb of any neuron is adjacent to the soma of not more than one neuron (the same or another); the separating gap is a synapse. Each endbulb is either excitatory or inhibitory (not both).

We call the neurons (zero or more) on which no endbulbs impinge input neurons; the others, inner neurons. (McCulloch and Pitts say "peripheral afferent neurons" for the former, but it is convenient to have a shorter phrase. "Efferent neurons" might be used for the latter, but it is not clear to us that this is appropriate. (As the present paper is only a working paper, we welcome suggestions as to improvements in the terminology.)

At equally separated moments of time (which we take as the integers on a time scale, the same for all neurons in a given net), each neuron of the net is capable of firing or not firing (being quiet) in an all-or-nothing manner. For an input neuron, the firing or non-firing at any time t is determined by conditions outside the net. One can suppose each is impinged upon by a sensory receptor organ, which under suitable conditions in

the environment causes the neuron to fire at time t. For an inner neuron, the condition for firing at time t is that at least a certain number h (the threshold of that neuron) of the excitatory endbulbs, and none of the inhibitory endbulbs, synapsing on it belong to neurons which fired at time $t-1$.

For illustration, consider the following nerve net, with input neurons J, K, L, M, and N and inner neuron P. Excitatory endbulbs are shown as dots, and inhibitory as circles. The threshold of P is 3 as shown by the figure on the triangle representing its soma. The formula written below the net expresses



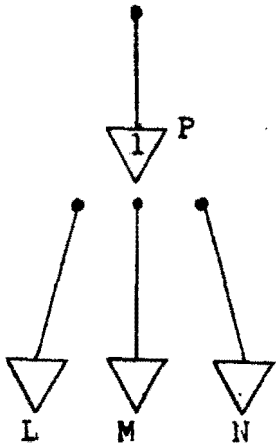
$$P(t) \equiv J(t-1) \,\&\, K(t-1) \,\&\, L(t-1) \,\&\, \overline{M(t-1)} \,\&\, \overline{N(t-1)}$$

Fig. 1

in logical symbolism that neuron P fires at time t [in the symbols, "$P(t)$"], if and only if (in symbols, "$\equiv$") all of J, K, and L and none of M and N fire at time $t-1$. ("&" means "and," and "$\overline{}$" means "not.").

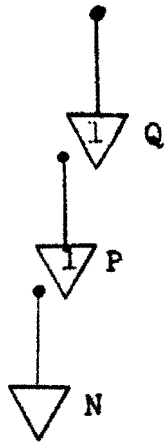
The method of nerve net construction illustrated in Fig. 1 applies for any number ≥ 1 of unnegated propositions (3 in Fig. 1) and any number ≥ 0 of negated propositions (2 in Fig. 1) combined conjunctively.

Two other nets (Figs. 2 and 3) illustrate additional methods which will be used in nerve net constructions in Sect. 5.



$$P(t) \equiv L(t-1) \vee M(t-1) \vee N(t-1)$$

Fig. 2



$$P(t) \equiv N(t-1).$$

$$Q(t) \equiv N(t-2).$$

Fig. 3

Here "$\vee$" means "or" (in the non-exclusive sense).

4. The Input to a Nerve Net: Consider a nerve net with k input neurons $N_1,\ldots,N_k$. The input (or experience) over all past time up to the present moment inclusive can be represented by a table or matrix with k columns corresponding to the input neurons, and with rows corresponding to the moments counting backward from the present moment $t = p$. The positions are filled with 0's and 1's, where 0 is to stand for quiescence, and 1 for firing, of the neuron in question at the time in question.

For example, with $k = 2$ the matrix might be as follows:

t	N_1	N_2
p	1	0
$p-1$	1	1
$p-2$	0	1
$p-3$	1	1
. . .		

Fig. 4

The 1 in the first row and first column means that N_1 fired at time p; the 0 in the first row and second column that N_2 did not fire at time p; the 1 in the second row and first column that N_1 fired at time $p-1$; etc.

If this table is extended down infinitely, we have a representation of the input, thought of as extending over all past time. The discussion whether we should think of past time as infinite will be left to the place where it becomes crucial (Sect. 6.1). For the purposes of Sect. 5 we need merely assume that it extends back in each case as far as the number of rows of the matrix being considered there.

By an event we mean any property of the input. Thus, any subclass of the class of all the possible tables represents an event, which occurs when the table describing the actual input belongs to this subclass. In coin tossing or dice throwing, examples of events are "heads" or "eleven" (as sum of the numbers of spots on the uppermost faces of the two dice). Here examples are: (1) N_1 fired at time p. (2) N_2 did not fire

at time p, and N_1 fired at time $p-1$. (3) One of N_1 and N_2 fired at time p. (4) N_1 and N_2 both fired at time p. (5) N_2 fired at some time. (6) N_2 fired at every time except p. Of these, the (present and) past described by the table in Fig. 4 constitutes an occurrence of events (1), (2), (3), and (5), but not of (4), while we need to know the rest of the table to see whether it constitutes an occurrence of (6).

5. Definite Events:

5.1. "Definite events" defined: We shall first restrict ourselves to events which refer to a fixed period of time, consisting of the γ (≥ 1) moments $p-\gamma+1,\ldots,p$ ending with the present. This means that in any table such as that of Fig. 4 we consider only the uppermost γ rows; e.g., with $\gamma = 3$:

t	N_1	N_2	
p	1	0	$N_1(p)$ & $\overline{N_2(p)}$
$p-1$	1	1	& $N_1(p-1)$ & $N_2(p-1)$
$p-2$	0	1	& $\overline{N_1(p-2)}$ & $N_2(p-2)$.

Fig. 5

The formula at the right expresses the same as is expressed by the table; i.e., it says that N_1 fires at time p ("$N_1(p)$") and ("&") N_2 does not fire at time p ("$\overline{N_2(p)}$"), and N_1 fires at time $p-1$ ("$N_1(p-1)$"), etc.

We call an event referring to just these γ moments definite of length (or duration) γ. With k input neurons, there are

exactly $k\gamma$ entries in a table describing the input for these moments. Therefore, there are exactly $2^{k\gamma}$ possible such tables. Therefore, there are exactly $2^{2^{k\gamma}}$ definite events of length γ since any particular event (of length γ with k input neurons) is obtained by saying which of the $2^{k\gamma}$ tables would constitute (if they represented the actual past) an occurrence of the event.

We call an event positive, if it only occurs when at least one input neuron fires during the period to which the event refers. There are $2^{2^{k\gamma}}-1$ definite positive events, since now we exclude as an occurrence of the event that past described by the table of all 0's.

5.2 Definite positive events:

Theorem 1. To each of the $2^{2^{k\gamma}}-1$ definite positive events of length γ (with k input neurons), there is a nerve net having an inner neuron which fires at time $p+2$, if and only if the event occurs during time $p-\gamma+1$ to p.

This theorem, except for the remark that the "lag" can be held to 2, is given by McCulloch and Pitts (1943).

Proof: To illustrate, say the event is one which occurs if and only if the pattern of firings over the past is represented either by the table of Fig. 5 or by the following table:

t	N_1	N_2	
p	1	0	$N_1(p) \,\&\, \overline{N_2(p)}$
$p-1$	1	0	$\&\, N_1(p-1) \,\&\, \overline{N_2(p-1)}$
$p-2$	1	0	$\&\, N_1(p-2) \,\&\, \overline{N_2(p-2)}$.

Fig. 6

That is, just these two (out of the $2^{2 \cdot 3} = 64$) tables are to constitute an occurrence of the event. The event is described by the following logical formula:

$$\left[N_1(p) \,\&\, \overline{N_2(p)} \,\&\, N_1(p-1) \,\&\, N_2(p-1) \,\&\, \overline{N_1(p-2)} \,\&\, N_2(p-2)\right]$$

$$\vee \left[N_1(p) \,\&\, \overline{N_2(p)} \,\&\, N_1(p-1) \,\&\, \overline{N_2(p-1)} \,\&\, N_1(p-2) \,\&\, \overline{N_2(p-2)}\right]$$

Figure 7

This formula is a "disjunction" having two "members" or "terms," each of which is a "conjunction" having six "members" or "factors." The two terms correspond to the tables of Figs. 5 and 6, respectively.

A nerve net which represents the event with lag 2 is constructed as follows:

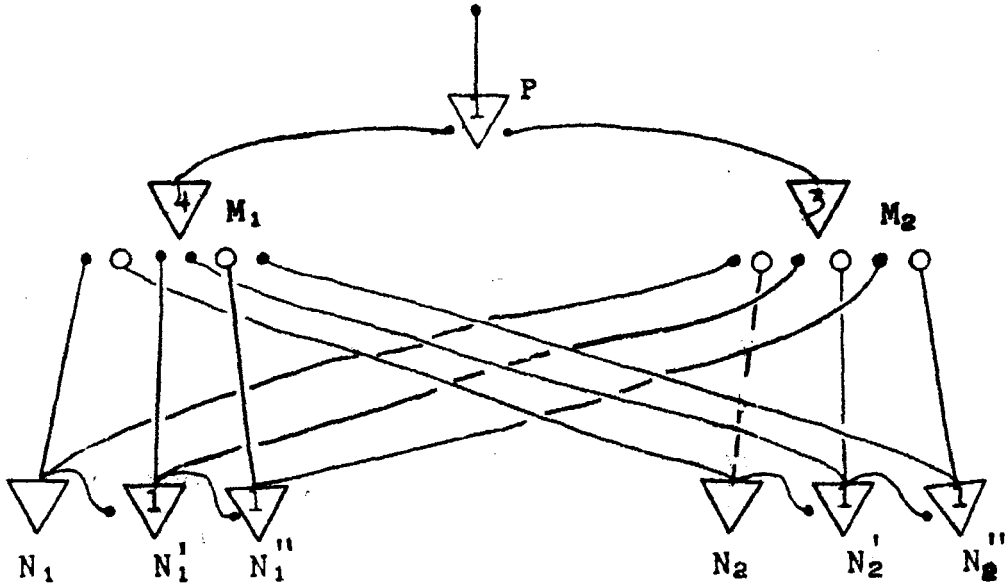


Figure 8

Using the method illustrated in Fig. 3,

$$\underline{N}_1'(p) \equiv \underline{N}_1(p-1), \qquad \underline{N}_2'(p) \equiv \underline{N}_2(p-1),$$

$$\underline{N}_1''(p) \equiv \underline{N}_1(p-2), \qquad \underline{N}_2''(p) \equiv \underline{N}_2(p-2).$$

Now using the device of Fig. 1,

$$\underline{M}_1(p+1) \equiv \underline{N}_1(p) \,\&\, \overline{\underline{N}_2(p)} \,\&\, \underline{N}_1'(p) \,\&\, \underline{N}_2'(p) \,\&\, \overline{\underline{N}_1''(p)} \,\&\, \underline{N}_2''(p)$$

$$\equiv \underline{N}_1(p) \,\&\, \overline{\underline{N}_2(p)} \,\&\, \underline{N}_1(p-1) \,\&\, \underline{N}_2(p-1) \,\&\, \overline{\underline{N}_1(p-2)} \,\&\, \underline{N}_2(p-2);$$

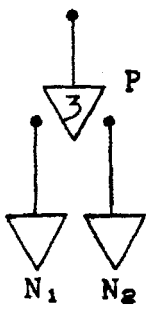
i.e., M_1 fires at time $p+1$ if and only if the past is described by the table of Fig. 5 (or the first conjunction in Fig. 7). Likewise, the firing of M_2 at $p+1$ corresponds to the table of Fig. 6 (or the second conjunction in Fig. 7). Finally, by the method of Fig. 2,

$$\underline{P}(p+2) \equiv \underline{M}_1(p+1) \vee \underline{M}_2(p+1).$$

Combining this with what has already been remarked, P fires at $p+2$ if and only if the event occurs during time $p-2$ to p.

The method of the illustration applies to every definite positive event which occurs for some one or more tables. By the restriction that the event be positive, each table must have at least one 1 in it, which assures the applicability of the device of Fig. 1.

There remains the case of the event which never occurs. This is represented, e.g., by the following net:



P never fires at time $p+2$
(or in symbols, e.g.,
$\underline{P}(p+2) \equiv \underline{N}_1(p) \,\&\, \overline{\underline{N}_1(p)}$).

Figure 9

5.3 Simpler nerve nets: While this proves the theorem, it is to be observed that often much simpler nets can be constructed than that given by the above method of proving the theorem.

Readers having technical acquaintance with symbolic logic will recognize that the construction used in proving the theorem corresponds to the principal disjunctive normal form of Hilbert-Ackermann (1928) which describes the event. In the illustration, the normal form is the formula of Fig. 7. Each of the tables which describe an occurrence of the event is represented by a conjunction or term in the normal form and is taken care of separately in building the nerve net. This makes the proof of the theorem simple, but the net complicated.

Consider for example the event which is described by saying that the table must be of one of the two following forms, where either a 0 or a 1 can be supplied independently for each blank "_."

t	N_1	N_2
p	—	0
$p-1$	—	—
$p-2$	—	1

or

t	N_1	N_2
p	—	—
$p-1$	—	1
$p-2$	—	—

$$\left[\overline{N_2(p)} \,\&\, N_2(p-2)\right] \vee N_2(p-1).$$

Figure 10

In terms of complete tables, this event would be expressed by a choice between $2^4+2^5-2^3=40$ tables, including that of Fig. 5 as one of them. The principal disjunctive normal form would be a disjunction of 40 conjunctions. The simple formula shown in Fig. 10 which represents it is a disjunctive normal form (not a principal one). The event is represented in the sense of the theorem by the following net.

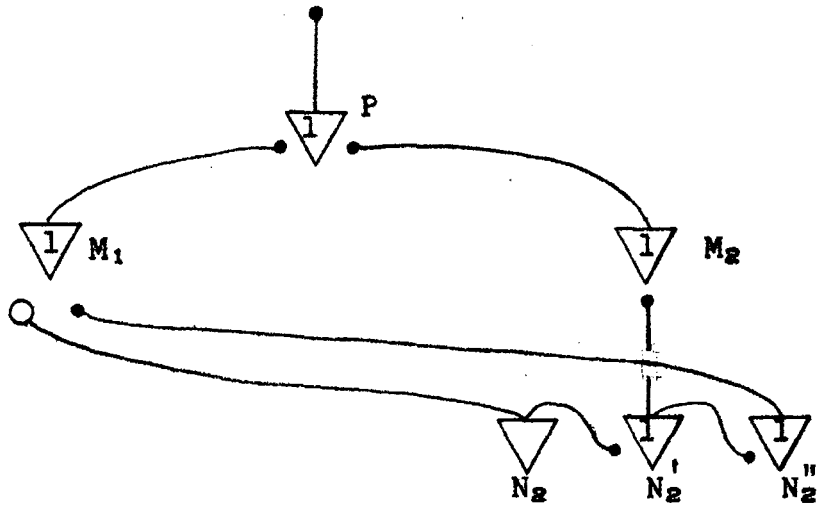
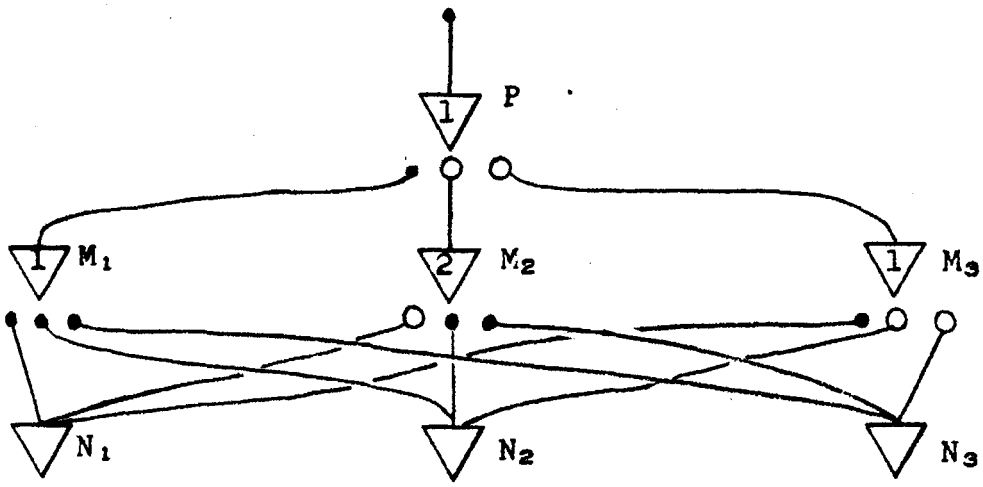


Figure 11

In this net we show only N_2 as an input neuron, although we defined our events in terms of two input neurons N_1 and N_2 in our illustrations. The net of Fig. 11 can constitute a part of a larger net having N_1 also as input neuron, entering in such a way that it has no endbulbs on any of the neurons shown in Fig. 11. The example illustrates that if we begin by defining events relative to a set $N_1,\ldots,N_k$ of input neurons, we need actually use in our net constructions only those of $N_1,\ldots,N_k$ whose firing or non-firing affects whether the event occurs.

There is a corresponding treatment, with the same lag, for conjunctive normal forms. We begin by considering the following

illustration, in which the normal form is a principal one with $\underline{k} = 3$ and $\underline{\ell} = 1$.



$$\underline{P}(\underline{p}+2) \equiv \left[\underline{N}_1(\underline{p}) \vee \underline{N}_2(\underline{p}) \vee \underline{N}_3(\underline{p})\right] \& \left[\underline{N}_1(\underline{p}) \vee \overline{\underline{N}_2(\underline{p})} \vee \overline{\underline{N}_3(\underline{p})}\right] \& \left[\overline{\underline{N}_1(\underline{p})} \vee \underline{N}_2(\underline{p}) \vee \underline{N}_3(\underline{p})\right] .$$

Figure 12

To see that this works, observe that we use Fig. 2 in obtaining M_1, so that

$$\underline{M}_1(\underline{p}+1) \equiv \underline{N}_1(\underline{p}) \vee \underline{N}_2(\underline{p}) \vee \underline{N}_3(\underline{p});$$

but Fig. 1 to obtain M_2 and M_3, so that

$$\underline{M}_2(\underline{p}+1) \equiv \overline{\underline{N}_1(\underline{p})} \& \underline{N}_2(\underline{p}) \& \underline{N}_3(\underline{p}), \quad \underline{M}_3(\underline{p}+1) \equiv \underline{N}_1(\underline{p}) \& \overline{\underline{N}_2(\underline{p})} \& \overline{\underline{N}_3(\underline{p})}.$$

Hence

$$\overline{\underline{M}_2(\underline{p}+1)} \equiv \underline{N}_1(\underline{p}) \vee \overline{\underline{N}_2(\underline{p})} \vee \overline{\underline{N}_3(\underline{p})}, \quad \overline{\underline{M}_3(\underline{p}+1)} \equiv \overline{\underline{N}_1(\underline{p})} \vee \underline{N}_2(\underline{p}) \vee \underline{N}_3(\underline{p}).$$

Also, we used Fig. 1 to obtain P, so that

$$\underline{P}(\underline{p}+2) \equiv \underline{M}_1(\underline{p}+1) \& \overline{\underline{M}_2(\underline{p}+1)} \& \overline{\underline{M}_3(\underline{p}+1)}.$$

Substituting our formulas for $\underline{M}_1(\underline{p}+1)$, $\overline{\underline{M}_2(\underline{p}+1)}$ and $\overline{\underline{M}_3(\underline{p}+1)}$ in the latter gives the formula for $\underline{P}(\underline{p}+2)$ in the figure.

This method of treating a principal conjunctive normal form depends on that disjunction which has no negated propositions' being one of the factors; but it must always be for a positive event, since otherwise the falsity of all the elementary propositions would make every term of the conjunction true.

If the principal disjunctive normal form has n terms, the principal conjunctive normal form has $2^{k\ell} - n$ factors, and vice versa (so the longer one form is, the shorter the other). We see why this is so in our illustration thus (omitting "(p)" after each "N" for brevity).

$$\begin{aligned}
&\left[N_1 \vee N_2 \vee N_3\right] \,\&\, \left[N_1 \vee \overline{N}_2 \vee \overline{N}_3\right] \,\&\, \left[\overline{N}_1 \vee N_2 \vee N_3\right] \\
&\equiv \overline{\left[\overline{N}_1 \,\&\, \overline{N}_2 \,\&\, \overline{N}_3\right] \vee \left[\overline{N}_1 \,\&\, N_2 \,\&\, N_3\right] \vee \left[N_1 \,\&\, \overline{N}_2 \,\&\, \overline{N}_3\right]} \\
&\equiv \left[N_1 \,\&\, N_2 \,\&\, N_3\right] \vee \left[N_1 \,\&\, N_2 \,\&\, \overline{N}_3\right] \vee \left[N_1 \,\&\, \overline{N}_2 \,\&\, N_3\right] \vee \left[\overline{N}_1 \,\&\, N_2 \,\&\, \overline{N}_3\right] \vee \left[\overline{N}_1 \,\&\, \overline{N}_2 \,\&\, N_3\right]
\end{aligned}$$

Under the bar in the second expression we have the principal disjunctive normal form of the negation of the first expression, so the last expression (which is the principal disjunctive normal form of the original expression) is obtained by combining disjunctively those 5 of the 8 elementary conjunctions which do not appear in the second expression.

When an event can be represented by a conjunctive normal form other than the principal one, a corresponding simplification can be made in the net construction just as in the case of disjunctive normal forms.

By using a normal form of either kind, we have held the lag s in the representation to 2.

Now it may happen that the most compact formula we have at hand to represent a given definite positive event is not a normal form. Then we can construct a nerve net of exactly corresponding structure and complexity, if we accept a greater lag s. In fact, the lag will be exactly the "depth" (or number of "layers") in the formula in terms of the operations $\&$ and $\vee$. We shall see this in Sect. 5.4 (Theorem 2).

For some events, of course, a lag of 1 suffices (or even a lag of 0 or -1 or -2, etc., if respectively the event specifies nothing about the firings at time p or times $p-1$, p or times $p-2$, $p-1$, p, etc. Reduction of the lag below 2 is not possible in general (with the assumed kind of neuron). A counterexample is the event $N_1(p) \,\&\, \overline{(N_2(p) \vee N_3(p))}$. To represent this with lag 1, the net would have to consist of the representing neuron P with endbulbs belonging directly to N_1, N_2, and N_3. One readily sees that no such net represents the event in question.

To hold the lag to 2 in all cases by use of a normal form, we may be obliged to have a very large number of endbulbs synapsing on a given soma, or of axons emerging from a given neuron. Biologically there are limitations. A relatively small increase in the lag will cut these numbers down. For example, a soma with 10^6 excitatory endbulbs synapsing on it is replaceable with an increase of only 2 in the lag by a net made up so that only 10^2 endbulbs synapse on each soma (but, of course, now a large number of neurons are necessary).

5.4 Definite events in general:

Corollary. To each of the $2^{2^{k\gamma}-1}$ definite non-positive events of duration γ with k input neurons, there is a nerve net having an inner neuron which does not fire at time $p+2$, if and only if the event occurs in time $p-\gamma+1$ to p.

Proof. Denote the event by E, and by $\bar{E}$ the complementary event or negation of E, which occurs exactly if E does not occur. The set of tables, one of which the past must fit if $\bar{E}$ occurs, is the complement (in the set of all $2^{k\gamma}$ k by γ tables) of the set, one of which the past must fit if E occurs.

Now $\bar{E}$ is positive, so by the theorem (Sect. 5.2), there are a net and neuron which represent $\bar{E}$ by firing that neuron at $p+2$, and therefore represent E by not firing the neuron at $p+2$.

Theorem 2. Consider any logical expression E in terms of $\&$, $\vee$, $^-$ and propositions $N_i(t)$ $(1 \leq i \leq k,\ p-\gamma+1 \leq t \leq p)$ describing a definite event E of length γ with k input neurons. Then there is a nerve net of corresponding structure which represents E by firing or by not firing, according as E is positive or non-positive, a certain neuron at time $p+s$, where s is the depth of E in terms of $\&$ and $\vee$ only.

Proof: It will be convenient to assume there are no double negations in E, as can be arranged by use of the law of double negation $\bar{\bar{R}} \equiv R$. (This does not change the depth.)

First we give the treatment for the least depth 1. For convenience we take γ to be 1, writing "N_1," "N_2," etc., for "$N_1(p)$," "$N_2(p)$." But for $\gamma > 1$ we would merely need to use the

method of Fig. 3 to introduce neurons whose firing at $t = p$ represents the firing of the various input neurons at the earlier times.

By an elementary conjunction (elementary disjunction), we mean a conjunction of one or more factors (terms) each of which is an elementary proposition N_1, N_2, etc., or a negated elementary proposition $\overline{N}_1$, $\overline{N}_2$, etc. (By allowing one factor or term, a single proposition or negated proposition can be considered as either a conjunction or a disjunction here.)

Now we have four basic cases to treat.

Case 1: An elementary conjunction containing at least one unnegated factor, e.g., $N_1 \,\&\, N_2 \,\&\, N_3 \,\&\, \overline{N}_4 \,\&\, \overline{N}_5$. The event is then positive; so we want to represent it by the firing of a neuron at time $p+1$. Use Fig. 1 to obtain this neuron.

Case 2: An elementary conjunction containing only negated factors, e.g., $\overline{N}_1 \,\&\, \overline{N}_2 \,\&\, \overline{N}_3$. The event is non-positive. But now its negation $\overline{\overline{N}_1 \,\&\, \overline{N}_2 \,\&\, \overline{N}_3}$ is positive. The latter is equivalent to $N_1 \vee N_2 \vee N_3$. Use the method of Fig. 2 to represent this by a neuron firing at $p+1$; this neuron then represents the original event by non-firing at $p+1$, as we wished to have it represented.

Case 3: An elementary disjunction containing at least one negated term, e.g., $\overline{N}_1 \vee \overline{N}_2 \vee \overline{N}_3 \vee N_4 \vee N_5$. The event is non-positive. But its negation $\overline{\overline{N}_1 \vee \overline{N}_2 \vee \overline{N}_3 \vee N_4 \vee N_5}$ is positive, and the latter is equivalent to $N_1 \,\&\, N_2 \,\&\, N_3 \,\&\, \overline{N}_4 \,\&\, \overline{N}_5$. Use Fig. 1 to represent the latter by firing at $p+1$; then the original event is represented by non-firing at $p+1$.

Case 4: An elementary disjunction containing only unnegated terms, e.g., $\underline{N}_1 \vee \underline{N}_2 \vee \underline{N}_3$. The event is positive. Use the method of Fig. 2 to represent it by firing at $p+1$.

The cases are mutually exclusive, except that a single unnegated proposition $\underline{N}$ can be considered as under either Case 1 or Case 4, and a single negated proposition $\overline{\underline{N}}$ as under either Case 2 or Case 3. But for one input only (which must be unnegated for Fig. 1), Figs. 1 and 2 coincide; so the treatment is actually the same. (Indeed, for $\underline{N}$ or $\overline{\underline{N}}$ it is only to have an inner neuron which represents them that any treatment is necessary; otherwise, we could consider them as representing themselves at time p.)

The treatment of a formula with depth > 1 requires only iteration of the processes used in the four basic cases.

It will suffice to illustrate by a complicated example, in which the depth is 4:

$$\left\{ \left[(\underline{N}_1 \vee \underline{N}_2) \,\&\, \underline{N}_3 \,\&\, \overline{\underline{N}}_4 \right] \vee \overline{\left[\overline{\underline{N}}_5 \,\&\, (\underline{N}_6 \vee \overline{\underline{N}}_7) \right]} \right\} \,\&\, \overline{(\underline{N}_8 \vee \underline{N}_9)}.$$

4 ———————————————————————————————

3 ———————————————————————

2 ——————————— ———————————

1 ——— ——— ———

(For handy reference we took an example with all $\underline{N}$'s different, but they could be identified in any combinations.) The underlines indicate the parts of various depths. Also $\underline{N}_3$, $\overline{\underline{N}}_4$ and $\overline{\underline{N}}_5$ are parts of depth 1, which we could treat as "degenerate" elementary disjunctions, but there is no need to consider them thus here.

First we apply Case 4 to obtain a neuron M_1 which represents $\underline{N}_1 \vee \underline{N}_2$ by firing at $\underline{p}+1$, and one M_3 which represents $\underline{N}_8 \vee \underline{N}_9$ likewise; and Case 3 to obtain one M_2 which represents $\underline{N}_6 \vee \overline{\underline{N}}_7$ by non-firing at $\underline{p}+1$.

Next consider the two parts of depth 2:

$$(\underline{N}_1 \vee \underline{N}_2) \,\&\, \underline{N}_3 \,\&\, \overline{\underline{N}}_4, \qquad\qquad \overline{\underline{N}}_5 \,\&\, (\underline{N}_6 \vee \overline{\underline{N}}_7).$$

Replacing $\underline{N}_1 \vee \underline{N}_2$ (i.e., $\underline{N}_1(\underline{p}) \vee \underline{N}_2(\underline{p})$) by its equivalent $\underline{M}_1(\underline{p}+1)$, and $\underline{N}_6 \vee \overline{\underline{N}}_7$ by its equivalent $\overline{\underline{M}_2(\underline{p}+1)}$ (since it is the non-firing of M_2 at $\underline{p}+1$ which represents it), these become, respectively, in terms of inputs from M_1 and M_2 at $\underline{p}+1$ and of N_3, N_4 and N_5 at $\underline{p}$:

$$\underline{M}_1(\underline{p}+1) \,\&\, \underline{N}_3(\underline{p}) \,\&\, \overline{\underline{N}_4(\underline{p})}, \qquad\qquad \overline{\underline{N}_5(\underline{p})} \,\&\, \overline{\underline{M}_2(\underline{p}+1)}.$$

(As the inputs are no longer all at the same time $\underline{p}$, we show the times.) The left one we treat by Case 1 to obtain a neuron L_1 which represents it by firing at $\underline{p}+2$, and the right one by Case 2 to obtain one L_2 which represents it by non-firing at $\underline{p}+2$.

Next consider:

$$\left[(\underline{N}_1 \vee \underline{N}_2) \,\&\, \underline{N}_3 \,\&\, \overline{\underline{N}}_4\right] \vee \overline{\left[\overline{\underline{N}}_5 \,\&\, (\underline{N}_6 \vee \overline{\underline{N}}_7)\right]}\ .$$

Replacing the first term by its equivalent $\underline{L}_1(\underline{p}+2)$, and the second by its equivalent $\underline{L}_2(\underline{p}+2)$ (since $\overline{\underline{N}}_5 \,\&\, (\underline{N}_6 \vee \overline{\underline{N}}_7)$ is equivalent to $\overline{\underline{L}_2(\underline{p}+2)}$), we obtain

$$\underline{L}_1(\underline{p}+2) \vee \underline{L}_2(\underline{p}+2).$$

This we treat by Case 4 to obtain a neuron Q which represents it by firing at $\underline{p}+3$.

Finally, consider the entire expression. Replacing the two factors by their respective equivalents, we obtain

$$\underline{Q}(p+3) \,\&\, \overline{\underline{M}_3(p+1)},$$

which we treat by Case 1 to obtain a neuron P which represents it by firing at $p+4$.

Incidentally, we have discovered in the process that the event is positive--we did not need to take the trouble of settling which it was at the beginning.

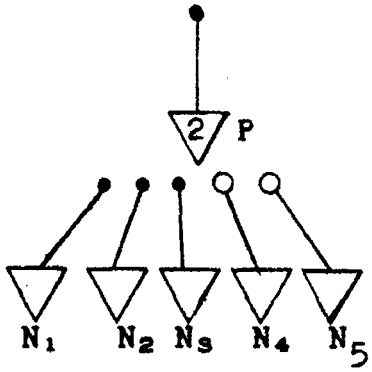
Both Theorem 1 and its corollary are corollaries of the present theorem, and the nerve net constructions in Sects. 5.2 and 5.3 for disjunctive and conjunctive normal forms are by the present method; so we might have given Theorem 2 first.

Other logical operations which might be used in defining events are definable in terms of $\&$, $\vee$ and $\overline{\ }$; e.g., $(\underline{F} \rightarrow \underline{G}) \equiv \overline{\underline{F}} \vee \underline{G}$ ("$\rightarrow$" is read "implies" or "if ..., then ..."), and $(\underline{F} \equiv \underline{G}) \equiv ((\underline{F} \rightarrow \underline{G}) \,\&\, (\underline{G} \rightarrow \underline{F}))$.

Summarizing, given any description, in words capable of being translated into logical symbolism, of a definite event, we have the means for constructing a nerve net to represent it of exactly corresponding complexity. So the theory of nerve net construction for definite events is as practical as one could ask. The lag can always be held to 2 for a given event, but sometimes a greater lag will correspond to a simpler description of the event, and give us a simpler net.

There may in special cases be simpler nets than those given by the method of proof of Theorem 2. We see this by considering

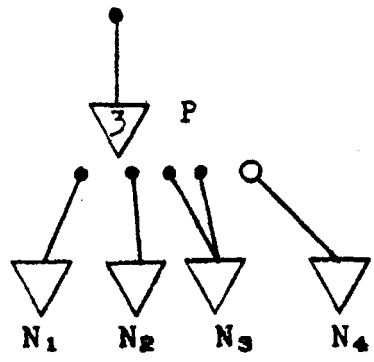
the condition for firing of any inner neuron of thereshold h at time t, which is that some set of neurons having a number $\geq h$ of endbulbs on it, and none of the neurons having inhibitory endbulbs on it, fire at time $t-1$. The condition for not firing is dual to this. For example, the nets below represent with lag only 1 the events described below them using expressions of greater depth (the upper by firing, the lower by non-firing):



$$\left[(N_1 \,\&\, N_2) \vee (N_1 \,\&\, N_3) \vee (N_2 \,\&\, N_3)\right] \,\&\, \overline{N}_4 \,\&\, \overline{N}_5$$

$$\left[\overline{N}_1 \vee \overline{N}_2\right) \,\&\, (\overline{N}_1 \vee \overline{N}_3) \,\&\, (\overline{N}_2 \vee \overline{N}_3)\right] \vee N_4 \vee N_5$$

Figure 13



$$(N_1 \vee N_2) \,\&\, N_3 \,\&\, \overline{N}_4$$

$$(\overline{N}_1 \,\&\, \overline{N}_2) \vee \overline{N}_3 \vee N_4$$

Figure 14

(Compare Fig. 14 with the treatment of $(N_1 \vee N_2) \,\&\, N_3 \,\&\, \overline{N}_4$ in the long example for Theorem 2.) We have not undertaken to study how much net simplification might be gained by attempting to use this method systematically with the help of appropriate logical transformations.

5.5 Representation of events in general: We can now prove the remark we made in Sect. 1 that there is no loss of generality in considering the representation of an event to consist of the firing or the non-firing (as appropriate) of a single neuron.

By saying that an event (occurring over a time ending with the moment p) is represented in a nerve net at a certain time $p+s$ ($s \geq 0$), we mean that some property of the state of the net at time $p+s$ is equivalent to the event having occurred ending at time p; i.e., according as the event did or did not occur, the net will or will not have that property.

But what happened at times $< p$ can only affect the state of the net at times $p+s$ for $s \geq 0$ via the state of the net at time p.

Say besides the k input neurons there are m inner neurons. The state of the net at time p consists of the condition (firing or non-firing) of each of the $m+k$ neurons. Thus, there are exactly 2^{m+k} possible states at time p. There are $2^{2^{m+k}}$ properties of the state of the net at time p. Any event ending at time p which can be represented at time $p+s$ is thus equivalent to one of these $2^{2^{m+k}}$ properties of the state at time p.

But for each of these, by applying the method of proof of Theorem 1 or its corollary, or of Theorem 2, to all the $m+k$ neurons (instead of only the input ones) and to only the moment p (instead of the interval $p-\chi+1$ to p), we can add additional neurons to get a neuron P which will fire or not fire (according as the property of the $m+k$ neurons at time p is positive or not) at time $p+2$, if and only if the $m+k$ neurons fulfill the property at time p; and hence, if and only if the event in question (referring to input neurons and ending at time p) occurred.

Incidentally, we have not made any assumption here whether the event in question is definite or not.

In organizing a complex of stimuli into a complex of responses (Sect. 1) as economically as possible, it is to be expected that the representation of events will not always be compressed into the form of the firing or non-firing of a single neuron.

5.6 Nerve nets without circles: A circle (of length c) in a nerve net is a set of distinct neurons $N_1,\ldots,N_c$ such that N_i has an endbulb on N_{i+1} ($i = 1,\ldots,c-1$) and N_c has one on N_1.

Theorem 3: Given any nerve net without circles and any inner neuron N in that net, the firing (non-firing) of that neuron at time $p+1$ is equivalent to the occurrence of a definite positive (non-positive) event ending at time p.

This theorem is stated for positive events by McCulloch and Pitts (1943).

Proof. Whether N fires at $p+1$ is completely determined by the firing or non-firing at p of those neurons $N_1',\ldots,N_r'$ having endbulbs on N. Consider those of $N_1',\ldots,N_r'$ which are inner neurons, and repeat the argument. Since there are no circles, any chain of neurons each impinged upon by an endbulb of the preceding must terminate. Let $\chi+1$= the length of any longest such chain; a longest must exist since there are finitely many such chains, and $\chi \geq 1$ since N is inner. Then the process terminates after χ steps. Thus, the firing or non-firing of N at $p+1$ is completely determined by the firing or non-firing of certain input neurons at times $p-\chi+1$ to p; i.e., it is equivalent to a definite event of duration χ. The event must be

positive, as firing can only be propagated but not originated under the law for an inner neuron's firing (Sect. 3).

Then of course N's non-firing at time $p+1$ is equivalent to the complementary event, which is non-positive.

(Any definite event is expressible by a logical formula, e.g., by a principal disjunctive normal form as in Sect. 5.2. So *a priori* there is a formula. By utilizing the condition for firing at each synapse, which we formulated in words in the last paragraph of Sect. 5.4 and could have in symbols, one can, of course, build up a formula in χ stages, as McCulloch-Pitts indicate.)

Corollary: *For a net without circles, any event ending at time p which can be represented by the firing (non-firing) of a given inner neuron N at a certain time $p+s$ ($s \geq 2$) is definite and positive (definite and non-positive).*

Proof: For by the theorem, the condition for the firing of N at time $p+s$ is the occurrence of a definite positive event ending at time $p+s-1$. But since by hopothesis, N's firing represents an event ending at time p, the input over time $p+1,\ldots,p+s-1$ cannot affect whether the aforesaid definite positive event occurs ending at time $p+s-1$. So in fact that definite positive event can be taken to refer only to a time ending at p.

This corollary constitutes the converse of Theorem 1 and corollary (or Theorem 2). Likewise, any event ending at p represented by a state, or a property of the state, at a time $p+s$ of a net ($s>o$) without circles, is definite.

6. Indefinite Events--Preliminaries:

6.1 Some examples: Consider the following nerve net (with a circle of length 1 consisting of M).

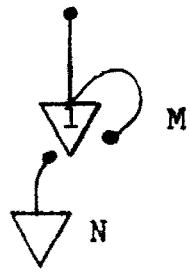


Figure 15

If at some time $t \leq p$ the neuron N fires, then the firing of M at time $p+1$ (and at every subsequent time) will follow. In symbols,

$$(Et)_{t \leq p} N(p) \longrightarrow M(p+1)$$

("$(Et)_{t \leq p}$" is read "there exists a $t \leq p$"). But we do not have equivalence ("$\equiv$") instead of merely implication ("$\longrightarrow$"), if past time is taken as infinite, since the firing of M at time $p+1$ can also be explained by firing of M at every past moment, without N having ever fired.

Similarly, the net

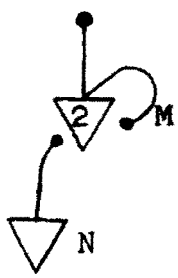


Figure 16

will only fire at time $p+1$ if N has fired at all past times;

in symbols

$$(\underline{t})_{\underline{t} \leq \underline{p}} \underline{N}(\underline{p}) \longleftarrow \underline{M}(\underline{p}+1)$$

(read "$(\underline{t})_{\underline{t} \leq \underline{p}}$" as "for all $\underline{t} \leq \underline{p}$"); but not conversely, for M may fail to fire at time $\underline{p}+1$ when $(\underline{t})_{\underline{t} \leq \underline{p}} \underline{N}(\underline{p})$ is true, by failing to fire over all past time.

$(\underline{Et})_{\underline{t} \leq \underline{p}} \underline{N}(\underline{p})$ and $(\underline{t})_{\underline{t} \leq \underline{p}} \underline{N}(\underline{p})$ are simple examples of events not referring to a definite period of past time; and we see that, under the assumption that past time is infinite, the nets shown do not represent them, by firing at time $\underline{p}+1$, in the sense of equivalence (the first is represented in the sense of "necessity" only, the second "sufficiency" only).

If we attempt to represent the former by non-firing, we have a net

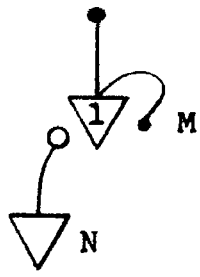


Figure 17

for which $(\underline{Et})_{\underline{t} \leq \underline{p}} \underline{N}(\underline{p}) \longrightarrow \overline{\underline{M}(\underline{p}+1)}$, but not conversely.

The difficulties encountered in these three examples are not escapable by using other nets to represent the events, or in other examples of indefinite events, but constitute the general rule for indefinite events. We shall show in Appendix 1 (Theorem 7, to be read after Part II) that, under the assumption of an infinite past, an event can be represented (in the sense of equivalence) by the firing or by the non-firing of a certain

inner neuron at time $p+s$ (for any fixed $s \geq 1$), only if the event is definite. For any net and inner neuron and s, and any indefinite event, it must either be possible to have a past for which the event does not occur and the neuron fires, or one for which the event does occur and the neuron does not fire, or both.

Of course, any living organism or actually constructed robot has only a finite past. The mentioned result shows that now we must take this into account; otherwise, we might have been tempted to use the fiction of an infinite past to simplify the theory.

6.2 Initiation: Accordingly, let us assume that the past for our nerve nets goes back from p (the present) a certain finite time only, the first moment of which shall be 1 on our time scale. (We find it more convenient notationally to call the first moment $t = 1$ than $t = 0$, but if we think of each positive integer t as referring to the final instant of a unit interval, this does make time start from 0.)

In seeking to represent events, we shall now assume the right not only to construct the nerve nets as we please, but also to fix the state (firing or non-firing) of each inner neuron at time 1. That is, we study representation of events in nerve nets started with a given internal state at the initial moment 1.

The range of the time variables in our logical formulas shall now be the integers from 1 forward, and this shall be the only part of the past we talk about except when we make it plain we intend otherwise.

Now the nerve net of Fig. 15, started at $\underline{t} = 1$ with M quiet, represents by the firing of M at $\underline{p}+1$ the event $(\underline{Et})_{\underline{t} \leq \underline{p}}N(\underline{t})$; and the net of Fig. 16, started with M firing, represents likewise $(\underline{t})_{\underline{t} \leq \underline{p}}N(\underline{t})$. That of Fig. 17 started with M firing represents $(\underline{Et})_{\underline{t} \leq \underline{p}}N(\underline{t})$ by the non-firing of M at $\underline{p}+1$. Thus, the two nets of Fig. 15 and 17 are able to remember if N has fired since their beginning by changing M from the state it had initially; while the net of Fig. 16 is able to recognize that N has never failed to fire by preserving M in the state it had originally, as Householder and Landahl (1945) have commented (p. 109).

To represent $(\underline{t})_{\underline{t} \leq \underline{p}}N(\underline{t})$ either by the firing or the non-firing of a neuron in a net with only N as input neuron, at least one inner neuron must be fired initially. For were all inner neurons quiet at time $\underline{t} = 2$, then in case $\overline{\underline{N}(1)}$ (i.e., if the input neuron N does not fire at time $\underline{t} = 1$), all neurons would be quiet at $\underline{t} = 2$. So the state of the inner neurons at $\underline{t} = 2$ would then be indistinguishable from that at $\underline{t} = 1$. Hence, the net at any time $\underline{p}+1 \geq 3$ would have the same state whether the past is

$\underline{t}$	1	2	3	4	...
$\underline{N}(\underline{t})$	0	1	1	1	...,

which makes $(\underline{t})_{\underline{t} \leq \underline{p}}N(\underline{t})$ false, or

$\underline{t}$	1	2	3	4	...
$\underline{N}(\underline{t})$	1	1	1	1	...,

which makes $(\underline{t})_{\underline{t} \leq \underline{p}}N(\underline{t})$ true.

The case in which all inner neurons are initially quiet is natural neurologically; the other case leaves it unexplained how the firing of certain inner neurons is to be produced at $\underline{t} = 1$. Of course, a natural explanation would be available by setting the time origin back, if the initial state in question is one which could be brought about by a suitable pattern of firings of the input neurons over a finite preceding time at the beginning of which all the inner neurons are quiet. But this is not so; e.g., in the case of the simple net for $(\underline{t})_{\underline{t}} \leq \underline{p}^{\underline{N}(\underline{t})}$ (Fig. 16).

Although $(\underline{t})_{\underline{t}} \leq \underline{p}^{\underline{N}(\underline{t})}$ cannot be expressed (for N as sole input neuron) without having an initially fired inner neuron, $(\underline{Eu})_{\underline{u}} \leq \underline{p}^{\left[\underline{K}(\underline{u}) \,\&\, (\underline{s})_{\underline{s}} < \underline{u}^{\overline{\underline{K}(\underline{s})}} \,\&\, (\underline{t})_{\underline{u}} \leq \underline{t} \leq \underline{p}^{\underline{N}(\underline{t})}\right]}$ with two input neurons K and N can be, by the following net, in which P fires at $\underline{p}+2$ if and only if the event occurs.

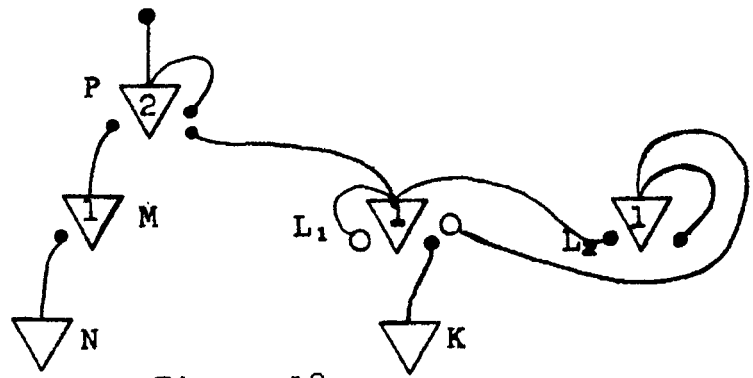


Figure 18

The neurons K, L_1 and L_2 act as a starting circuit, which can only be used once, for the generality circuit N, M and P.

The device with a modification is general. Suppose we have given a net in which an event is represented by the firing of a

certain inner neuron at time $p+s$ if certain inner neurons are initially fired.

First, let us add a starting circuit K, L_1 and L_2 from a new input neuron K, with axons from its neuron L_1 leading to all the same neurons and with the same kinds of respective endbulbs as the axons from each of the inner neurons which were initially fired in the given net.

Furthermore, each input neuron N of the original net we now make an inner neuron N' with a threshold of 2, and we insert new neurons N and R, the former taking over the role of the original N as input neuron, and the latter an inner neuron as shown. The heavy line stands for the axons which lead from the original N.

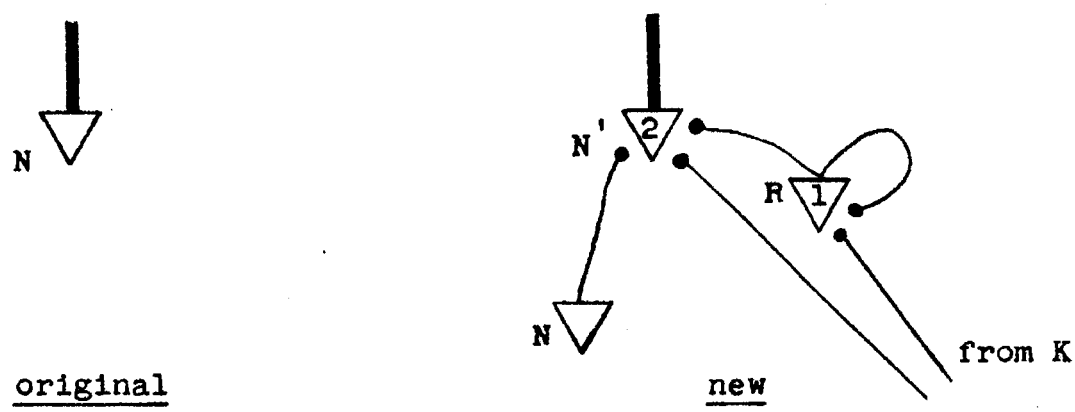


Figure 19

This accomplishes the double purpose of lagging the input from N by 1 and of blocking it for moments prior to the first moment u at which K is fired.

Now with all inner neurons initially quiet, no neuron except N can fire until the first moment u at which K fires.

Then at moment $u+1$ the neuron L_1 takes over the role of the original initially fired inner neurons, while N' carries the input of N at u. At every subsequent moment $u+v+1$ $(v \geq 1)$, all of the original neurons (counting N' as the original N) will behave as they formerly would have at time $v+1$, if the present input over $u,\ldots,u+v$ had been the input over time $1,\ldots,v+1$.

So the output neuron will fire at $p+s+1$, if and only if the event now occurs relative to $t = u$ instead of to $t = 1$; i.e., we have a representation of the event redefined to refer not to the whole past but to the past beginning with $t = u$, and with an increase of 1 in the lag in the representation.

Now if it is assumed that there are conditions in the environment which would continually stimulate K to fire, or that at least such a condition exists at $t = 1$, then our net will represent the event relative again to the whole past (since now $u = 1$). Thus, we are provided with a "natural" way of getting a representation of any event, referring to the input neurons besides K, which could be represented "unnaturally" by the firing of a neuron in a net started with some initially fired input neurons.

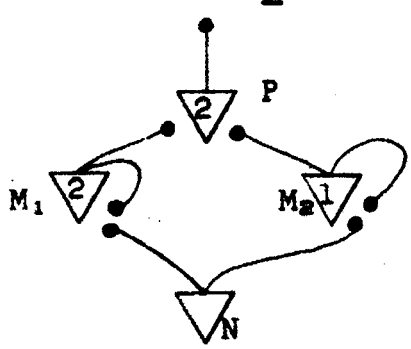
Here "natural" means only that we do not need to go outside the McCulloch–Pitts laws of neural behavior to fire some inner neurons at $t = 1$; but the starting circuit K, L_1, L_2, and the blocking circuit K, R, N' are not thought of as plausible mechanisms biologically. However, our first aim is to see what is at all possible, and one can then seek other and perhaps more

natural ways for accomplishing the same.

This argument that by assuming an initially stimulated input neuron K we can avoid having to have any initially fired inner neurons applies only to representation of events by firing a neuron, if for representation by non-firing at $p+s+1$ one wishes that the output neuron fire at times $1,\ldots,s+1$. The question involved will be analyzed in Sect. 6.3.

The same construction but omitting the K and L_2 of the starting circuit and the delaying-blocking neurons R and N', and firing L_1 initially, shows that it is always possible, if we are to use initially stimulated inner neurons, to hold the number of them to 1, without any increase in the lag. This again is for the case of representation by firing. For representation by non-firing, the situation is slightly more complicated, and we shall not go into it.

As stated, we ordinarily consider nets only for a specified initial state of the inner neurons. However, McCulloch and Pitts consider the problem of "solving" nets with their initial condition unspecified. To "solve" for a given inner neuron P, say at time $p+1$, means then to find for what inputs over time $1,\ldots,p$, and what initial states of the inner neurons, P will fire at $t = p+1$. Now in the following net, the necessary and sufficient condition that P fire at $p+1$ is that N fire at all times $< p$ and both M_1 and M_2 fire at time $t = 1$.



$$P(p+1) \equiv$$

$$(t)_{t<p} N(t) \,\&\, M_1(1) \,\&\, M_2(1).$$

Figure 20

This seems to be counterexample to the formula next after (9) on p. 126 of McCulloch-Pitts (1943), the proof of which we did not follow; for if we understand the formula correctly, it implies that the condition for firing should only require the existence of one (suitably chosen) neuron known to fire initially. In this example we cannot conclude that P fires at $p+1$ on the basis of any information which tells us only that one of the neurons fires at $t = 1$. (Our 1 seems to be their 0.) This apparent counterexample has discouraged us from further attempts to decipher Part III of the McCulloch-Pitts 1943.

6.3 Definite events reconsidered: Now that we have introduced the assumption that the past for a nerve net is finite, we must reexamine the treatment of definite events which was given in Sect. 5.

What happens now when $p < \gamma$; i.e., when the period of time to which the event is supposed to refer extends back to before the moment $t = 1$?

Generally, one may suppose that the durations γ of definite events which are significant for an organism will be small in relation to the age p of the organism at which the event is significant.

This, however, does not enable us to dismiss the problem. For to make the theory of nerve net control accurate, we should, for each definite event considered, either (a) show that an "hallucination" that the event has occurred arising during the first $\gamma-1+s$ moments of life could not have any serious effect

on behavior, or (b) provide against the occurrence of such an hallucination in the first $\chi-1+s$ moments.

When we use definite events to build indefinite ones, e.g., the event consisting of a certain definite event having occurred ending at some time $q \leq p$ (briefly, the memory of the definite event having occurred), such an hallucination could conceivably have a long-term effect, even if it has no immediate effect on behavior.

The solution by (a) is, of course, outside the present investigation, and belongs rather to the full problem of organizing stimuli into responses (Sect. 1).

For organisms, the picture of the nervous system coming into activity in toto at a fixed moment $t = 1$ is implausible. But this means only that organisms (at least those which survive) do solve the problem for their process of coming into activity.

For machines, it is familiar that starting difficulties may have to be taken into account by the engineer.

To take a fictitious illustration, consider the case of the "rat satellite robots" for the Tuvian Navy. A rat satellite robot is intended to go about a ship, and whenever after three hours (= $\chi-1$ moments) it has not smelled a rat, and at the next moment (the χ-th) land is in sight, the robot abandons ship.

The robots were ordered from RAND and were built by the Robotry Section from blueprints prepared by the Logicians Group on the basis of the theory in Sect. 5 above, with two input

neurons, N_1 which is fired by the smell of a rat, and N_2, which is fired by the sight of land. The inner neuron P, which fires at time $p+2$ if the event "no rat smelled for $\chi-1$ moments, land seen at the χ-th moment" has occurred during the time $p-\chi+1$ to p, was connected to an effector mechanism for abandoning ship at time $p+3$.

Suitable ceremonies were scheduled for the occasion of their installation in the harbor of the Tuvian Naval Base.

When the occasion arrived, they were placed on board the ships, and their batteries were connected up supplying power for operating the nerve nets and effector mechanisms. But three moments later, just as the Tuvian Grand Admiral was congratulating the RAND delegate, all the robots went overboard!

Proceeding to details, it is, of course, a matter of definition how we shall interpret "events of duration χ" when $p < \chi$. But whatever definition is adopted, we must keep the facts about nerve net behavior straight.

We shall (as best suits our present purpose, which is to lay a firm basis for the theory in Sect. 7) say that an event E of duration χ can only have occured ending at p when $p \geq \chi$.

Then, of course, the logical formulas we have used to represent the events in Sect. 5 are not complete. If E_1 is the formula which described a definite event of length χ there, the formula E which describes it fully now is $E_1 \& p \geq \chi$. The negation E of this is $E_1 \vee p < \chi$, while the formula for the "complementary" event of duration χ is rather $E_1 \& p \geq \chi$. Thus, some care is now necessary in connection with the operation of

negation. The theory in Sect. 5 is applicable to the part of the formula which does not give the time reference; i.e., the relationships studied there apply to the $\underline{E}_i$ and $\underline{\overline{E}}_i$.

Now consider a nerve net as constructed for Theorem 1 in Sect. 5.2 to represent an event occupying the time interval $p-\gamma+1$ to p of length γ by firing a certain inner neuron (the "output neuron") at time $p+2$.

Using this net under the restriction now that the life of the net starts with a certain moment $t = 1$, and under the stipulation that at that moment all inner neurons are quiet, it is clear that the output neuron will fire at any time $p+2$ for $p \geq \gamma$ correctly; i.e., if and only if the event occurrrd in the time $p-\gamma+1$ to p.

But the net might also fire at a time $< \gamma+2$; namely, this could happen if and only if the event is such that our present initial condition of the inner neurons (all quiet) is one which could also take place in Sect. 5.2 at some moment m where $p-\gamma+2 \leq m \leq p+1$ for some occurrence of the event in $p-\gamma+1$ to p.

So assume (in the context of Sect. 5.2) that we have an occurrence of the event in the course of which all inner neurons are quiet at $t = m$.

The state of the inner neurons at time m would then have to be the same (i.e., all quiet) if the table describing the past is altered to show only 0's for all input neurons at all times $t < m$. For from a past consisting entirely of non-firings prior to m, no firing of any inner neuron can be produced at time m.

Since the state of the inner neurons is unchanged at time m, and the inputs for $t = m,\ldots,p$ are unchanged, the output neuron still fires at $t = p+2$. So by Theorem 1, the event still does occur in time $p-\gamma+1$ to p; i.e., we now have an instance of the event occurring in which no input neurons were fired in $p-\gamma+1$ to $m-1$; in particular (since $m \geq p-\gamma+2$), the event can occur with input 0 on all its input neurons at its first moment $p-\gamma+1$.

Conversely, if this is the case, the output neuron will fire now at time $\gamma+1$, if in times $1,\ldots,\gamma-1$ the inputs are what they could be in Sect. 5.2 for the moments $p-\gamma+2$ to p of an occurrence of the event with only 0's for $p-\gamma+1$.

Call a definite event of length γ prepositive if the event can occur only when some input neuron fires in its first moment $p-\gamma+1$; i.e., the selections from among the $2^{k\gamma}$ possible $k \times \gamma$ tables which describe occurrences of the event all have at least one 1 in their bottom row. The prepositive events are a subclass of the positive events.

Now we have shown that a necessary and sufficient condition that no "hallucination" be possible (in the sense of the output neuron's firing at a time t when the event has not occurred ending at time $t-2$) is that the event be prepositive.

We gave the reasoning for the case in which the event is to be represented by firing of a neuron at time $p+2$ (corresponding to Theorem 1), but it applies equally well to the other cases in Sect. 5; i.e., to representation by firing at $p+s$ for any given $s \geq 1$, or to representation by non-firing at $p+s$

(then hallucinations are always possible, since the event is necessarily non-positive by Theorem 3 or its corollary), or to representations by a property of the state at a certain time $p+s$ ($s \geq 0$). (In giving the sufficiency proof, we write the inequality on m now $p-\gamma+2 \leq m \leq p+s$, and change the input for $t \leq \max(m-1,p)$.)

Most but not all events we may wish to consider will be prepositive.

The analysis is valid for any net which operates correctly when $p \geq \gamma$, whether constructed as in Sect. 5 or not, and started now with all inner neurons quiet.

If a positive event of length γ is not prepositive, we can build a net which represents it by firing a neuron P at time $p+2$ (or $p+s$ for some $s \geq 1$), if this net is started at $t = 1$ with one of its inner neurons fired (but all others quiet), as follows. We simply take an inner neuron L as in Figure 21, initially fired (as the "+" indicates).

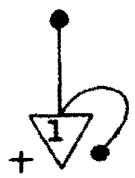


Figure 21

We then treat this as though it were an additional input neuron, required to fire at $t = p-\gamma+1$, in applying the method of net construction of Sect. 5. (Of course, then more than one axon may be required from L to other neurons.)

This device also makes it possible to represent non-positive events of length $\underline{\gamma}$ by firing of a neuron at time $\underline{p}+2$ (or $\underline{p}+\underline{s}$ for some $\underline{s} \geq 1$); in Sect. 5 they would have to be represented by non-firing.

Another device for fixing a net, constructed as in Sect. 5 to represent a positive but not prepositive event (then $\underline{\gamma} \geq 2$), so that no hallucination can be produced, is to let the inhibitory endbulb of L_1 in the following net impinge upon the output neuron of P of the net of Sect. 5. The number of the L's is $\underline{\gamma}+\underline{s}-2$.

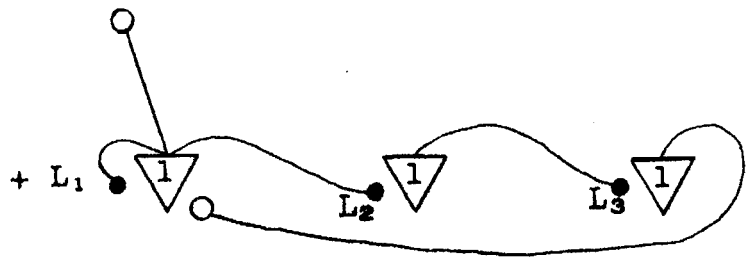


Figure 22

(Drawn for $\underline{\gamma}+\underline{s} = 5$.)

We can also use this to fix nets constructed as in Sect. 5 to represent a non-positive event by non-firing at $\underline{p}+\underline{s}$, if we change the endbulb of L_1 which is to impinge on P to a set of excitatory endbulbs equal in number to the threshold of P, and also fire P itself at time 1. (If $\underline{\gamma}+\underline{s} = 2$, no L's are added.)

The devices of Figs. 21 and 22 seem artificial, and not likely to be found in organisms. We point them out to save the need for making an exception of non-prepositive events in the theory. If mechanical realizations of McCulloch-Pitts neurons are used in controlling robots, such devices might be useful.

The upshot of the analysis is that only by reference to artificially produced firing in inner neurons at $\underline{t} = 1$ could an organism recognize complete absence of stimulation of a given duration, not preceded by stimulation; it would not know whether the stimulation had been absent, or whether it had itself meanwhile come into existence.

If instead of the initially fired inner neuron of Fig. 21 we use an input neuron subject to continual environmental stimulation, then all events can be taken to be prepositive by referring them to the class of input neurons as enlarged to include K.

This is plausible biologically, if we also grant that the mathematical model is probably too exact in that it gives too much emphasis to a single neuron at a single moment of time (.0005 sec.). It is unlikely that any such input at a single moment would by itself result in any significant overt action or memory.

Having chosen to investigate a precise model, it is not to be expected that all aspects of this model will be equally pertinent to the reality from which the model is abstracted.

6.4 Why consider indefinite events? Since the lifetime of an organism or machine is always finite, having an end as well as a beginning, why is it not sufficient to consider only definite events?

The number of moments (identifying a moment with a synaptic delay of .0005 sec.) in a human lifetime of 100 years is of the

order of 3×10^{12}.

To construct a nerve net, treating events as definite, that would account for behavior at 60 years of age influenced by stimuli at 10 years, we would need chains of neurons of length 1.5×10^{12}. If the event were at all complicated, we would need large numbers of such chains. Moreover, we would need further mechanism to provide for this same behavior occurring at 61 years or 59 years due to stimuli at 10 years, or indeed for each value of $\underline{d}$ where $\underline{d}$ ranges from the smallest elapsed time after 10 years at which the behavior can be influenced up to the greatest, and is measured in units of .0005 sec. We do not necessarily need a whole new set for each value of $\underline{d}$, since many neurons can be made to serve in common for various values of $\underline{d}$, e.g., the delay chains for various values of $\underline{d}$ greater than a given one $\underline{d}_i$ could have their first $\underline{d}_i$ neurons in common. But at least each intermediate value of $\underline{d}$ would, up to the greatest in question, require some structural additions, new axons if not new neurons.

All this would have to be duplicated for every sort of event which occurring at one time could influence behavior at all later times in life.

The total number of neurons is only of the order of 10^{18}.

To use definite events as a mathematical basis for explaining human behavior in all its flexibility over a lifetime of 3×10^{12} moments thus appears altogether unrealistic.

To emphasize what is meant, take the case of Solomon Grundy. On the afternoon of Monday he burns his hand on the

stove. Then one nerve net tells him not to touch the stove on Tuesday, a different one (at least in part) on Wednesday, and so on.

If he outlives the life expectancy for which his delay chains are designed, he must thereafter suffer an advancing amnesia; for each day added beyond his expectancy at the end of life he completely forgets one day at the beginning.

Humans and animals do not function in this way, though simple mechanisms for learning and subsequent forgetting in robots could be devised on this basis.

Indeed, calculations on the amount of information recorded in the memory (cf. McCulloch 1949) make it difficult to explain memory entirely in terms of McCulloch-Pitts neurons on any basis, a fortiori, certainly not in such an uneconomical way as by setting up only nets for definite events. So it is necessary (if perhaps in the end it will not be sufficient) to go beyond the present stage of our analysis.

It thus appears that the appropriate mathematical abstraction for us now is to treat the problem of explaining behavior as though organisms and machines were immortal, having an infinite future though a finite past. We want to provide for behavior that could be used ad infinitum, if merely the nerve net and effector mechanisms were immortal.

By trying to provide for behavior over an infinity of time by a finite mechanism, we have a model for the real problem of providing for complex behavior over a long finite

lifetime by a relatively small mechanism.

The questions of reducibility of other mechanisms to McCulloch–Pitts nerve nets (not always without increasing the size of the mechanisms) is significant on this basis, but trivial on the basis of explaining behavior over a fixed finite time only.

7. Regular Events:

7.1 "Regular events" defined: We shall presently describe a class of events which we will call "regular events." (We would welcome any suggestions as to a more descriptive term.*)

We assume for the purpose that the events refer to the inputs up through time p on a set of k input neurons $N_1,\ldots,N_k$ the same for all events considered; but the definition applies equally well for any $k \geq 1$ or even for $k = 0$.

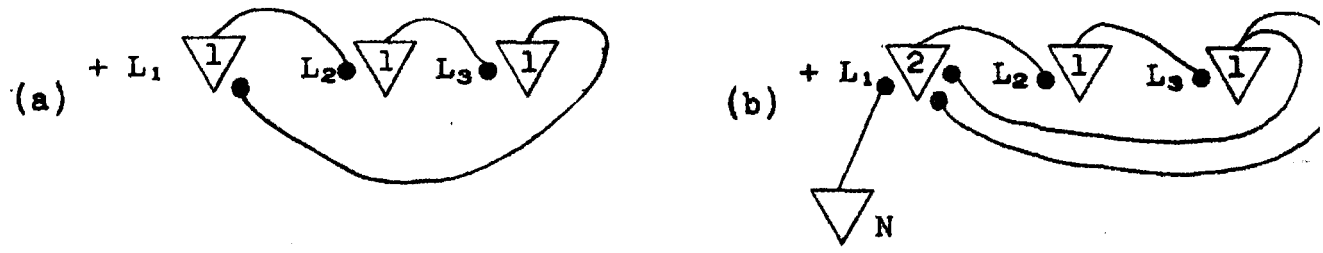
The events can refer to the value of p. Our objective is to show that all and only regular events can be represented by nerve nets or finite automata. We have already seen in Sect. 6.3 why reference to the time is called for; but it may be illuminating to consider some examples from the point of view of solving given nets.

Consider first the net of Fig. 22 taken by itself (the inhibitory endbulb from L_1 is superfluous now). The condition for L_1's firing (under the assumption that it is fired initially), i.e., the event represented by L_1's firing, is given by formula

$$L_1(p) \equiv p \leq 3.$$

---

*McCulloch and Pitts use a term "prehensible," introduced rather differently; but since we did not understand their definition, we are hesitant to adopt the term.



$L_1(p) \equiv p = 1 \bmod 3.$

Figure 23

In defining "regular events " we shall build on the notion of a definite event of length γ, as originally introduced in Sect. 5.1 and completed by adding & $p \geq \gamma$ to the definition in Sect. 6.3.

But now we extend the class of definite events by providing that the description of such an event may also contain the specification that the first moment is 1; i.e., that $p-\gamma+1 = 1$, i.e., that $p = \gamma$. Events with this specification we call initial. So now there are $2^{2^{k\gamma}+1}$ definite events of length γ with k input neurons, namely the $2^{2^{k\gamma}}$ as before (non-initial) and the $2^{2^{k\gamma}}$ initial ones.

Now each event we shall build up, starting from (and including) the $2^{2^{k\gamma}+1}$ definite events for each γ, will be interpretable in the following way. The statement that the event has occurred (ending at time p) is equivalent to the statement that one of a non-empty finite or infinite class of definite events has occurred (ending at time p). (More precisely, the class may be infinite if the value of p is unknown, but for a fixed value

of $\underline{p}$ there are, of course, only finitely many events which could have occurred.)

Our class of regular events will be defined inductively, starting with the definite events, and using three operations $E \vee F$, EF, E^*F; i.e., it shall be the least class containing the definite events, and closed under these operations.

Given any events E and F already built up from definite events by zero or more applications of the operations, by the event $E \vee F$ we shall mean the event which occurs if E occurs or F occurs. In other words, the class of definite events which can constitute an occurrence of $E \vee F$ is the sum of the respective classes for E and F.

Clearly the operation is associative; i.e., $(E \vee F) \vee G \equiv E \vee (F \vee G)$; so the parentheses can be omitted. The reason for writing equivalence here with four bars will appear in Sect. 7.3.

For example, if E and F are definite events of durations $\underline{\ell}$ and $\underline{m}$, respectively, then $E \vee F$ is an event which occurs exactly when an event (of length $\underline{\ell}$) belonging to E occurs, or one (of length $\underline{m}$) belonging to F occurs, or both. One might be tempted to regard this as a definite event of duration $\max(\underline{\ell},\underline{m})$; but this would be wrong, since supposing $\underline{\ell} > \underline{m}$ it could occur when $\underline{p} < \underline{\ell} = \max(\underline{\ell},\underline{m})$, namely by F occurring. Also, this would not give what we want when either of the next two operations is applied.

Given any events E and F already built up by the operations, by EF we shall mean the event which consists in E having just occurred preceded immediately by F having occurred. Thus, EF

has occurred ending at time p, exactly when one of the class of events which can constitute an occurrence of E (say this one is of length χ) has occurred in time $p-\chi+1$ to p, and one of those which can constitute an occurrence of F (say this one is of length m) has occurred in the period $p-\chi-m+1$ to $p-\chi$. (Note: We have chosen our notation EF so that we proceed back into the past in reading from left to right.) This operation is also associative.

For example, if E, F, and G are definite events of lengths χ, m, and n, respectively, any occurrence of $(E \vee F)G$ will be a definite event of one of the lengths $\chi+n$ or $m+n$. (Say $\chi > m$. By refraining above from interpreting $E \vee F$ as of duration χ, now when $(E \vee F)G$ occurs by F occurring ending at time p, the preceding occurrence of G must end at time $p-m$, not $p-\chi$.)

If E is an initial definite event, and F is any event of our class, then EF is an event which never occurs; since for E to occur, its first moment must be $t = 1$; so under the interpretation in Sect. 6.3, EF_1 is impossible for any one of the definite events F_1, F_2,... whose occurrence can constitute an occurrence of F. Thus, in this case EF is represented by the firing of P in the net of Fig. 9.

If E and F are events already constructed, then by E*F we shall mean the event which consists of zero or more consecutive occurrences of E preceded by one of F. That is, E*F can occur whenever

$$\overbrace{E \ldots E}^{n \text{ times}}F$$

occurs for some $n \geq 0$.

The reason we do not define E* separately as a unary operation (expressing that E has occurred consecutively zero or more times) instead of E*F as a binary operation, is that then for $\underline{n} = 0$ an occurrence of E* would be of duration 0; but (at least for convenience) we are requiring the lengths of our definite events to be always ≥ 1.

To say that E has happened one or more times we can write E*E.

For example, if E, F, and G are definite events of lengths $\underline{\gamma}$, $\underline{m}$, and $\underline{n}$, respectively, an occurrence of $(E \vee F)^*G$ must be of a definite event of one of the lengths $\underline{a}\underline{\gamma} + \underline{b}\underline{m} + \underline{n}$ $(\underline{a}, \underline{b} \geq 0)$.

We reflect now that we have two systems of notation for events: (A) logical notations for definite events as used in Sect. 5 (with the addition of $\underline{p} \geq \underline{\gamma}$ in Sect. 6.3 and $\underline{p} = \underline{\gamma}$ above) and for some other events in Sects. 6.1 ff.; (B) our newly introduced notations for regular events starting with single capital letters as representing definite events.

There will be ambiguity if we use (A) as the starting point for (B) instead of capital letters, unless we are careful to show the durations $\underline{\gamma}_1, \ldots, \underline{\gamma}_s$ of each of the definite events $E_1, \ldots, E_s$ used as the units for the construction of the regular events.

The question of translatability between the systems of notation (A) and (B) in either direction has not yet been examined thoroughly.

However, we have verified that the notations (B) can be translated into notations (A) with, of course, a sufficient amount of mathematical apparatus added to the logical notations. The details are technical and are given in Appendix 2.

But it may be instructive now to give a few simple examples of translation in the other direction, i.e., from (A) to (B). The conventions regarding parentheses are those of algebra with $E \vee F$, EF, and $E*F$ analogous to $e+f$, ef, and e^2f. Also, the associative law $(E*F)G \equiv E*(FG)$ permits omitting parentheses, as well as the two associative laws already mentioned.

The event of duration γ which happens for all inputs over the interval $p-\gamma+1$ to p we call the identical event of length γ; for $\gamma = 1$ we write it as I, then in general I^{γ} ($\equiv$ I ... I to γ factors).

Let the result of adding $p = 1$ to the specifications for a definite event E, to make it initial, be written E^0.

For any event E of length 1 the negation $\bar{E}$ is also definite of length 1.

For the present illustrations, let the event of length 1 that N fires at time p (in symbols, $N(p)$ be written simply N; that K fires simply K; and that both K and N fire be written L.

Now the events described as follows in the left column are expressed by the corresponding notations in the right column.

(See next page)

$(\underline{Et})_{\underline{t}} \leq \underline{p}^{\underline{N}(\underline{t})}$	$I*N$
$(\underline{t})_{\underline{t}} \leq \underline{p}^{\underline{N}(\underline{t})}$	$N*N^{0}$
$(\underline{Eu})_{\underline{u}} \leq \underline{p}^{\left[\underline{K}(\underline{u}) \,\&\, (\underline{t})_{\underline{u}} < \underline{t} \leq \underline{p}^{\underline{N}(\underline{t})}\right]}$	$N*K$
$(\underline{Eu})_{\underline{u}} \leq \underline{p}^{\left[\underline{K}(\underline{u}) \,\&\, (\underline{s})_{\underline{s}} < \underline{u}^{\overline{\underline{K}(\underline{s})}} \,\&\, (\underline{t})_{\underline{u}} \leq \underline{t} \leq \underline{p}^{\underline{N}(\underline{t})}\right]}$	$N*L^{0} \vee N*L\bar{K}*\bar{K}^{0}$
$\underline{N}(\underline{t})$ for at least two values of $\underline{t} \leq \underline{p}$	$I*NI*N$
$\underline{N}(\underline{t})$ for exactly one value of $\underline{t} \leq \underline{p}$	$\bar{N}*N^{0} \vee \bar{N}*N\bar{N}*\bar{N}^{0}$, call this M
$\underline{N}(\underline{t})$ for an odd number of values of $\underline{t} \leq \underline{p}$	$(\bar{N}*N\bar{N}*N)*M$
$\underline{p} \geq 3$	I^{3}
$\underline{p} = 1$	I^{0}
$\underline{p} = 1 \bmod 3$	$(I^{3})*I^{0}$
$\underline{p} \leq 3$	$I^{0} \vee II^{0} \vee I^{2}I^{0}$

7.2 An algebraic transformation: We list several equivalences:

(1)	$(E \vee F) \vee G$	$\equiv$	$E \vee (F \vee G)$.	Associative laws
(2)	$(EF)G$	$\equiv$	$E(FG)$.	
(3)	$(E*F)G$	$\equiv$	$E*(FG)$.	
(4)	$(E \vee F)G$	$\equiv$	$EG \vee FG$.	Distributive laws
(5)	$E(F \vee G)$	$\equiv$	$EF \vee EG$.	
(6)	$E*(F \vee G)$	$\equiv$	$E*F \vee E*G$.	
(7)	$E*F$	$\equiv$	$F \vee E*(EF)$.	
(8)	$E*F$	$\equiv$	$F \vee E(E*F)$.	

Under the definition just given, each regular event is obtained by building it up from certain definite events as the units by zero or more applications of the three operations. Of course, these constructions are by no means unique.

Lemma 1. For any $s \geq 2$: Every regular event can be expressed as a finite disjunction of one or more regular events, each of which is either definite of length $< s$ or is an event constructed out of units each of length $\geq s$. (Also true trivially for $s = 1$.)

Of course, we can always understand there is at most one of the latter, since any disjunction of them is again one.

We write out the proof for the case $s = 2$.

The lemma is true when the given event is definite; then there is just one term in the disjunction, which is of the first or the second kind according as its length is 1 or more.

Likewise, if E and F each have the property described in the lemma, so does $E \vee F$.

Now say E and F are as described, and consider EF. By use of the distributive laws (4) and (5), EF is equivalent to a disjunction of terms, each of which has one of the following forms

$$E^1F^1, \qquad E^{(2)}F^{(2)}, \qquad E^{(2)}F^1, \qquad E^1F^{(2)},$$

where in each case ^{1} indicates a definite term of length 1, and $^{(2)}$ indicates a term composed out of units each of length ≥ 2.

Now a term E^1F^1 can be construed as a definite event of length 2; so it is of the second kind for the theorem upon considering it as one of the units.

Also $E^{(2)}F^{(2)}$ is of the second kind.

Now consider $E^{(2)}F^1$. Using (2), (3), and (4), the F^1 can be moved progressively inward until finally F^1 occurs only in parts of the form HF^1 where H is definite and of length ≥ 2. Then each such part can be taken as a unit, which will be of length ≥ 3.

For the last form $E^1F^{(2)}$, we proceed similarly using (2) (from right to left), (5) and (8) (in combination with (5) and (2)).

Now say E and F are as described in the theorem, and consider E*F. By use of (6) we can then get E*F equivalent to a disjunction of terms of the two forms $E*F^1$ and $E*F^{(2)}$. For illustration (noting the remark just following the theorem), say, e.g., E is $E_1^1 \vee E_2^1 \vee E^{(2)}$. So we have now two possibilities,

$$(E_1^1 \vee E_2^1 \vee E^{(2)})*F^1, \qquad (E_1^1 \vee E_2^1 \vee E^{(2)})*F^{(2)}.$$

Consider the former. This is an event of which an occurrence must consist of one occurrence of F^1 followed by $n \geq 0$ occurrences of various of the events E_1^1, E_2^1 and $E^{(2)}$. Let $G_1,\ldots,G_9$ be all products of two of E_1^1, E_2^1, and $E^{(2)}$; i.e., G_1 is $E_1^1E_1^1$, G_2 is $E_1^1E_2^1$, etc. Then an occurrence of the event is the same as an occurrence of one of F^1, $E_1^1F^1$, $E_2^1F^1$ or $E^{(2)}F^1$, followed by zero or more occurrences of any of $G_1,\ldots,G_9$. Thus, in symbols (and using (6) next and then (7)):

$$(E_1^1 \vee E_2^1 \vee E^{(2)})*F^1$$

$$\equiv (G_1 \vee \ldots \vee G_9)*(F^1 \vee E_1^1F^1 \vee E_2^1F^1 \vee E^{(2)}F^1)$$

$$\equiv (G_1 \vee \ldots \vee G_9)*F^1 \vee (G_1 \vee \ldots \vee G_9)*(E_1^1F^1 \vee E_2^1F^1 \vee E^{(2)}F^1)$$

$$\equiv F^1 \vee (G_1 \vee \ldots \vee G_9)^*(G_1 \vee \ldots \vee G_9)F^1 \vee (G_1 \vee \ldots \vee G_9)^* (E_1^1 F^1 \vee E_2^1 F^1 \vee E^{(2)} F^1).$$

Now each of $G_1, \ldots, G_9$ can be handled as was one of E^1F^1, $E^{(2)}F^{(2)}$, $E^{(2)}F^1$, $E^1F^{(2)}$ in the case for EF above; then $G_1 \vee \ldots \vee G_9$ is composed out of units of length ≥ 2. Then by the method for $E^{(2)}F^1$ in the case for EF above, $(G_1 \vee \ldots \vee G_9)F^1$ and $E^{(2)}F^1$ are likewise, while $E_1^1F^1$ and $E_2^1F^1$ are definite of length 2. So the entire expression obtained last is of the desired form. Like arguments apply to $(E_1^1 \vee E_2^1 \vee E^{(2)})^*F^{(2)}$. Finally, any disjunction of expressions of the desired form is of the desired form.

7.3 Identity and equivalence: In dealing with regular events, special care is necessary to distinguish between senses of "equivalence." As we introduced them, any regular event is identified with a class of definite events; and two regular have thus far been treated as equivalent only if these classes of definite events for the two are the same.

An event is a partition of all the possible inputs over the whole past for the nerve net into two classes, those inputs for which the event occurs, and those for which it does not occur.

What we have called a "non-initial definite event of length χ" is a partition of all the pasts for the net into two classes, such that all pasts of length $< \chi$, i.e., for which $p < \chi$, are in the second class (those for which the event does not occur), while those pasts of length $\geq \chi$ are in the first or the second class according as the input over the last χ

moments has or has not a certain property; i.e., the classification is independent of the input prior to $p-\gamma+1$.

But could two non-initial definite events which are distinct in the value of γ or the property over $p-\gamma+1$ to p be the same as events?

They could in one case, namely when the property over $p-\gamma+1$ to p is impossible of occurring; this is the case which was treated by Fig. 9 in Sect. 5.2. These definite events of length γ for various values of $\gamma \geq 1$ are all the same as events. We may call this event, which never occurs, the improper (or impossible) event.

Outside of this exception, an event can be a definite event in only one way. For suppose we have an example of an input over time 1 to p for which the event occurs. Then we may seek the least $\gamma \leq p$ such that the event also occurs when the value of p is changed to γ and the input over time 1 to γ ($= p$) is what it was formerly over time $p-\gamma+1$ to p. This γ must be the length of the event; and the property of the input over the last γ moments which defined the event is obtained by considering what inputs over this time give occurrences of the event.

Similar remarks apply to "initial definite events of length γ." Here all pasts for which $p \neq \gamma$ are in the second class. The initial definite events of length γ which never occur are all the same as events.

Combining the cases of non-initial and initial definite events, an event can be a definite event in only one way (i.e., either non-initial or initial, but not

both, with only one length $\underline{\ell}$, and with only one property of the input over $\underline{p}-\underline{\ell}+1$ to $\underline{p}$), except for the improper event which can be construed as either non-initial or initial and with any $\underline{\ell}$. So actually there are only $2^{2^{\underline{k}\underline{\ell}}+1}-1$ distinct definite events of length $\underline{\ell}$ with $\underline{k}$ input neurons; in Sect. 7.1 we counted the improper event twice, once as a non-initial and once as an initial event. For $\underline{k} = 0$, there are thus just 3 events of length $\underline{\ell}$, the possible non-initial one $I^{\underline{\ell}}$, the possible initial one $I^{\underline{\ell}-1}I^{0}$, and the impossible one $\overline{I}$.

Now consider a regular event of the form $E \vee F$, where E and F are definite of length $\underline{\ell}$. Quite evidently the E and F are not uniquely determined from the event. For example, there might be two $\underline{k}$ x $\underline{\ell}$ tables exactly for which E occurs, a third for which F occurs. By recombining, taking E_1 as occurring when the first table applies, and F_1 when the second or the third applies, we get the event as $E_1 \vee F_1$ with different components, or indeed, the event can be considered as one event of length $\underline{\ell}$.

Now, in fact, our transformations of events in Sect. 7.2 were such as to preserve the class of definite events underlying a given regular event, except that sometimes, e.g., E^1F^1 was reconstrued as a definite event of length 2. To make it exact what transformations shall be allowed, we can reconsider a regular event as given by saying which of various tables of length $\underline{\ell}$ for various $\underline{\ell}$, with or without specification that $p = \underline{\ell}$ (rather than merely $\underline{p} \geq \underline{\ell}$) would describe an occurrence of it. For if it is given by saying which of various definite

events would describe an occurrence of it, then we can replace each of these by the tables (zero or more up to $2^{k\gamma}$ of them) which constitute an occurrence of the respective definite event. We shall say that two regular events, as given in the notations of Sect. 7.1 starting with specified definite events, are identical if the resulting classes of k x γ tables (for various γ) are the same. We write identity by $\equiv$.

The empty class of k x γ tables goes with the improper event; call this event $\bar{I}$. We have:

(9) $E \vee \bar{I} \equiv \bar{I} \vee E \equiv E$.

(10) $E\bar{I} \equiv \bar{I}E \equiv \bar{I}$.

(11) $E^*\bar{I} \equiv \bar{I}$.

(12) $\bar{I}^*E \equiv E$.

These permit simplifications of events into which $\bar{I}$ is built; in fact, all $\bar{I}$'s can be removed, unless the whole becomes $\bar{I}$.

Now, unfortunately, given an event as simply a partition of the possible inputs over the whole past for the net, the class of k x γ tables in terms of which it can be constructed as a regular event is not unique.

Consider the example of $N \vee NI^*I$ and N, where $k = 1$, and N signifies the event of length 1 consisting of the firing of the one input neuron N at time p.

The only k x γ table for N is that having a 1 in its one position. But $N \vee NI^*I$ has this table, also both tables of length 2 agreeing in the first row, also all four tables of length 3 agreeing in the first row, etc.

But $N \vee NI^*I$ and N each occur, if and only if the input neuron fires at time p; so as events they are the same. We call this sameness equivalence, and write $N \vee N^*I \equiv N$.

The importance of the distinction is that from $E \equiv F$ we can infer $EG \equiv FG$, $GE \equiv GF$, $E^*G \equiv F^*G$, and $G^*E \equiv G^*F$; but we cannot make the first and third inferences in terms of equivalence $\equiv$. In particular, $N \vee NI^*I \equiv N$, but not $(N \vee NI^*I)N \equiv NN$. (Of course, $E \equiv F$ implies $E \equiv F$; but not conversely.) As another example, $I^*I \equiv I^*I^o$, but $I^*IN \not\equiv I^*I^oN \equiv T$ (by (13) below and (11)).

Summarizing, our theory of regular "events," with our operations $E \vee F$, EF and E^*F and the relation $\equiv$ apply to classes of $k \times l$ tables (fixed k and varying l) in terms of which we can represent the events, rather than to the events in the simple sense. More particularly, it is the two operations EF and E^*F for which the class of tables for E, rather than merely the resulting event E, must be known, because the lengths of the tables enter into the meaning of the operations.

It would thus be more explicit to say that we are dealing with a theory of certain expressions for events ("representations" would be a good word, if we were not using it already in another sense).

We now extend our notion of "prepositive" to initial events, by saying that all except the improper one (which is at the same time non-initial and as such prepositive under the definition in Sect. 6.3) are not prepositive. Single $k \times l$ tables are special cases of definite events of length l; so the definition applies to them.

Now we say that a regular event as given by a class of $k \times \gamma$ tables (fixed k and varying γ) is prepositive, if all the tables of the class are prepositive.

In Sect. 6.3 we saw that prepositiveness was necessary and sufficient that a nerve net with all inner neurons initially quiet constructed to represent a non-initial definite event of length γ when $p \geq \gamma$ should also represent it correctly (without "hallucinations") when $p < \gamma$.

The extension to initial definite events preserves this as a necessary and sufficient condition for representability with all inner neurons initially quiet; the necessity is clear by reasoning similar to that in Sect. 6.3, and the sufficiency holds because there is no such prepositive event except the impossible one.

Furthermore, now a sufficient condition that in representing an event it be possible to take all inner neurons quiet initially is that there be a way of expressing the event in terms of definite events and our three operations for which the corresponding class of tables (or of definite events) is prepositive. This will be included as part of the next theorem.

To get a necessary condition, we introduce the idea of a minimal set of $k \times \gamma$ tables (fixed k and varying γ) for an event. Start with any set of $k \times \gamma$ tables for the event, and to each table consider the least segment of it ending at time p for which all backward extensions describe occurrences of the event. Replace the table by this. Carrying out the process for each table in the given set, we get a minimal set.

The minimal set so obtained is unique for a given event, as one gets the same minimal set by first extending each given table to an initial table in all possible ways (which method gives the complete set of tables, which is unique for the event); and then minimizing this (by the above process which leads to a unique result), we get the same class of tables as in minimizing directly.

Of course, the method of minimizing is not described "constructively," and one question which arises at once is whether a constructive minimization process for a set of tables corresponding to a regular event as expressed in terms of definite events and our three operations exists.

Another question is whether the minimal set of tables must, for a regular event, necessarily be one which corresponds to an expression for the event in terms of definite events and the operations. (The complete table does, as will follow from the proof of Theorem 6 in Sect. 9.)

We do not go into these questions, which one would naturally investigate if the study is to be continued.

However, we can now say that a necessary condition that a regular event be representable by a net with all inner neurons initially quiet is that the minimal set of tables for it be prepositive.

Some algebraic simplifications are possible when initial definite events enter into an expression for a regular event. Say E^0 is an initial definite event. Then for any regular event F:

(13) $E^0F \equiv I.$

(14) $E0^*F \equiv I.$

Used along with (2) – (4) (and for simplification (9) – (12)), we come out with the result that no initial event need enter into an expression for a regular event, other than as an "earliest" event in the following sense.

For a given expression for a regular event in terms of definite events, we define recursively as follows which occurrences of definite events in it are earliest.

In a regular event given as simply a definite event, that definite event is earliest.

The earliest events in E and the earliest events in F are the earliest in E∨F.

The earliest events in F are the earliest in EF and in E*F.

7.4 Representability of regular events:

Theorem 4: To each regular event, a nerve net can be constructed which, when started in a prescribed way, represents the event by firing a certain inner neuron at time p +2, if and only if that event has occurred ending at time p inclusive. If the given event is prepositive, the representation can be by a net started with all inner neurons quiet.

Proof is based on Lemma 1, Sect. 7.2, for s = 2. The Theorem is true for $\bar{1}$, by Fig. 9, Sect. 5.2; and for other events, by Sect. 7.3, we can exclude $\bar{1}$ as a unit.

So first we give the proof for the case of an event (not $\bar{1}$) constructed out of units (not $\bar{1}$) each of length 2 or more.

This we do by induction on the number n of occurrences of units in the expression for the event.

In the induction we will arrange at each stage that the neuron which is to fire at time $p+2$ will be (as in Sect. 5.2, since $\bar{I}$ is excluded) one of threshold 1 impinged upon by only excitatory endbulbs (i.e., it effectuates a disjunction operation) with no axons feeding back into the net.

If $n = 1$, then the event is a definite one E. We have three cases. (a) E is prepositive, hence not initial. The net is as given in Sect. 5.2, the reasoning that this net works being supplemented as in Sect. 6.3. (b) E is not initial and not prepositive. We use the treatment given in Sect. 6.2 employing Fig. 21. (c) E is initial. Then we use an inner neuron as follows, treated for the net construction of Sect. 5.2 as though it were an additional input neuron required for the occurrence of the event to fire at time $p-\gamma+1$.

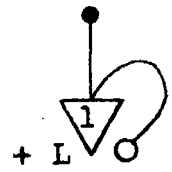


Figure 24

This, of course, is simply a neuron whose condition for firing is $p = 1$.

Now if $n > 1$, the event under consideration is of one of the forms $E \vee F$, EF, and E^*F where E and F are each constructed from $< n$ units.

First, suppose the event is $E \vee F$. Then by the hypothesis of the induction we can construct nets to represent E and to represent F, say with representing neurons P and Q, respectively, each with threshold 1 and only excitatory endbulbs impinging, and with no axons feeding back. To represent $E \vee F$ we "identify" P and Q; i.e., we replace them by a single neuron —call it P—having all the endbulbs which separately impinged on P and on Q, and we similarly identify the input neurons $N_1,\ldots,N_k$ for the two nets, i.e., the axons which led from $N_1,\ldots,N_k$ in the net for E, and those in the net for F now both lead from $N_1,\ldots,N_k$. The construction can be diagramed as follows, using heavy lines to represent a number of axons.

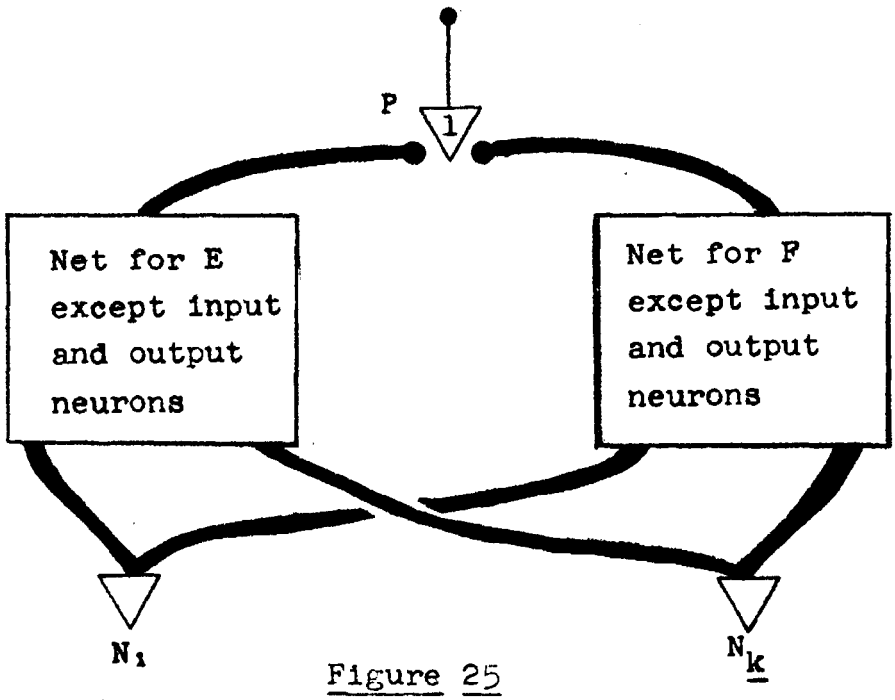


Figure 25

The heavy bundle of neurons leading to P from the left are those which would be required in the net for E separately; likewise from the right in the net for F. The bundle from N_1 toward

the left, those from N_1 in the net for E, to the right in the net for F, etc. The fact that the output neurons for the two given nets had no axons leading back in insures that they still operate independently of each other in this combination.

Next consider an event EF. In the construction out of units which we are using for E, consider those occurrences of units in it which are earliest. The events we are considering refer to k input neurons $N_1,\ldots,N_{\underline{k}}$. Now consider the event E' which is obtained from E by modifying each earliest unit to make it refer to one new neuron $N_{\underline{k}+1}$ which is required to fire at the second moment of each such earliest unit. There is such a second moment in the period of the unit, by our assumption in connection with the use of Lemma 1 that each unit is of length ≥ 2. Also, the resulting event E' is regular with the same number of occurrences of units, since this change in the earliest units only gives an event with the same structure in terms of its respective components by the operations $E \vee F$, EF, and E*F. So by the hypothesis of the induction on $\underline{n}$, we can represent this event E' by a net. However, we simplify the construction by leaving out the neuron of Fig. 21 in the case of earliest events in E' which come under Case (b) for definite events. (By remarks in Sect. 7.4, Case (c) can be excluded.)

Now the net for EF is obtained by identifying $N_{\underline{k}+1}$ in the net for E' with the output neuron Q of the net for F, and of course, identifying $N_1,\ldots,N_{\underline{k}}$ as input neurons for the two nets. The output neuron is that for E'. The construction can be

diagramed thus:

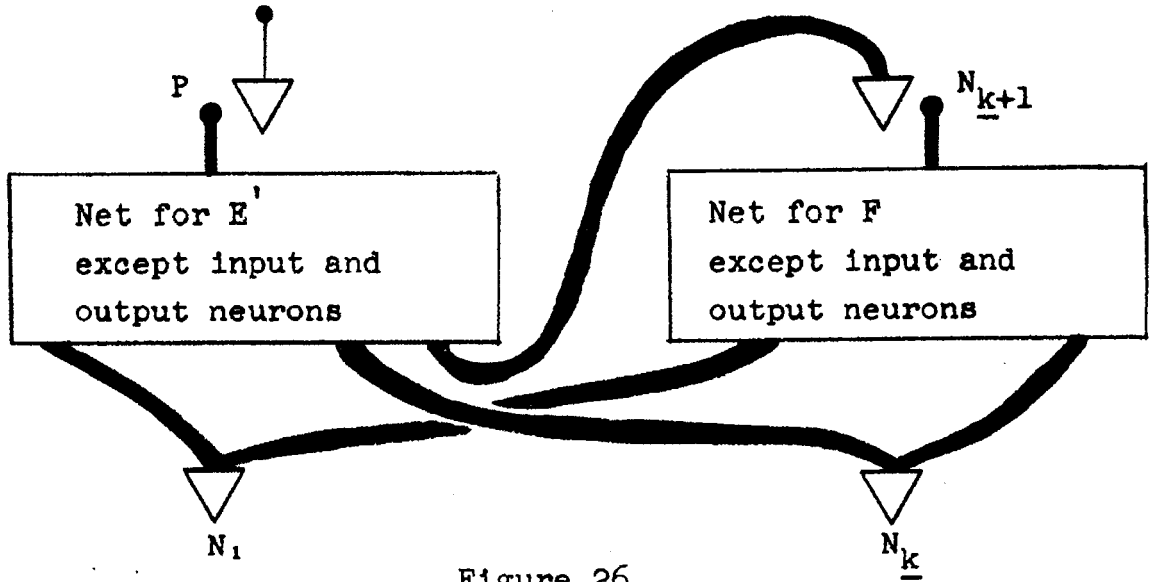


Figure 26

The event E' is positive, requiring a firing of N_{k+1} at its second moment. But N_{k+1} can be fired only at a time later than 2, since in its role of output neuron for the net for F it fires at time $p+2$ ($p \geq 1$) where p is the last moment of an occurrence of F. No "hallucination" is possible as a result of leaving out the neurons of Fig. 21 for the units in E' which were not prepositive, as this necessity that N_{k+1} fire at the second moment, which must be > 2, prevents. (In fact, the arguments of Sect. 6.3 that "hallucinations" can occur when an event is not prepositive do not apply now, since some inner neurons of the net for F will necessarily be firing at the first moment of these units of E'.) These remarks (with the avoidance of the neuron of Fig. 21) are necessary to establish the last remark of the theorem.

We have lastly the case for E*F. As in the preceding case we modify E to E'. Then we combine the nets obtained by

the hypothesis of the induction (omitting Fig. 21 in treating earliest units in E') as diagramed thus:

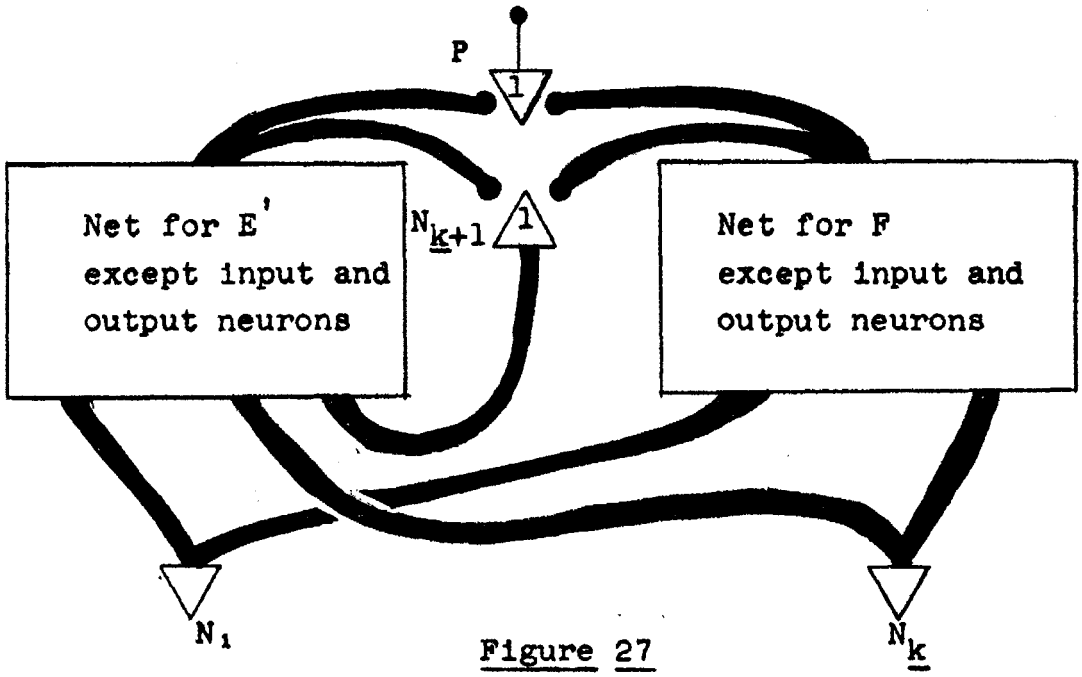


Figure 27

Under the assumption of infinite past time (as in Sect. 6.1), the firing of P could, of course, be explained by E having occurred repeatedly ad infinitum into the past. But here we are understanding that the whole net is started in a certain condition, which is either that all inner neurons are quiet or that some inner neuron (or neurons, if we prefer) in the net for F are fired, according as F (and therefore E*F) is prepositive or not. Then as in the reasoning under the treatment of EF, the net for E' can only be a cause of P and N_{k+1}'s firing if there has originally been a firing derived from the net for F, which serves as an input into that for E' that must be fired at the second moment for the latter. Of course, thereafter P will be fired on each repetition of E. (P and N_{k+1} must be separate, to meet the condition that the output neuron not feed back into the net.)

This completes the treatment for regular events constructed from units each of length ≥ 2.

Now by Lemma 1, and using the method already indicated for the basis (i.e., for $n = 1$) to treat definite events of length 1, and the method already used under the case of the induction step ($n > 1$) for $E \vee F$ to combine nets for different disjunctive members, we get the theorem for regular events in general.

7.5 Discussion of the proof and further problems: As we have already remarked in Sect. 6.2, the use of a net with initially fired inner neurons seems unnatural. But this is unavoidable, if we are to represent non-prepositive events, since we must (by examples such as are given at the beginning of Sect. 7.1) make our mathematical theory complete. A way of avoiding the use of such nets biologically, namely by considering only events dated from some environmental stimuli, has been indicated (Sect. 6.2).

A second respect in which the present proof seems artificial and leads to complicated nets is in the use of Lemma 1, the proof of which involves rather extensive reformulation of the events.

If we deal only with events which are already expressed in terms of units each of length ≥ 2, or have the form for Lemma 1, the proof of the theorem is straightforward and the nets constructed are simple, i.e., of a degree of complexity corresponding very well to the complexity of the given description of the event to be represented.

The necessity of using Lemma 1, if we do not restrict the events to be already expressed in the form for that lemma, arises from the fact that in Sect. 5 we needed a lag of 2 to represent definite events in general.

Some simple events can be represented with lag of 1; and for these the units into which we feed the outputs from the nets representing the preceding events would not need to be of length ≥ 2. Thus, to give a uniform treatment in proving the theorem, we resorted to a device (the proof of Lemma 1) or a restriction (that the property of Lemma 1 is already present) which can be dispensed with in many particular cases. This is why, e.g., in Sect. 6.2., simpler nets are available for representing certain indefinite events than would be given by the method of proof of the present theorem.

As was noted in Sect. 5.3, often definite events can be represented by simpler nets by using a lag greater than 2. Then the method of net construction for the proof of Theorem 4 would require the use of Lemma 1 for $\underline{s}$ = the greatest lag used in any of the units (or possibly not this great, depending on how the units enter). As the proof of Theorem 4 is given, also the net would have to be chosen so that the representing neuron appears as in Fig. 2; i.e., performs a disjunction operation. This could always be arranged by an increase of 1 in the lag. But also probably the proof can be adapted to apply directly, somewhat as the proof of Theorem 1 was generalized to get Theorem 2 (but we have not examined this in detail).

The difficulty which calls for Lemma 1 arises when we try to represent a repetition of some event which is shorter than the time necessary for the net to organize a representation of it by the firing of a single neuron; the solution by Lemma 1 consists in considering grosser events before attempting to represent repetitions of them.

If we consider that one or two synaptic delays are probably not significant for determining behavior in an organism (as we remarked at the end of Sect. 6.3), it seems that the complication is connected with an over-refinement in our model of the biological reality.

So we can urge that the methods of net construction used in proving the theorem are simple enough, granting that from the general method we can often start out to find simpler nets in special cases.

The question may occur to the reader, why did we select the particular three operations $E \vee F$, EF, and E^*F? When we say that the net constructions are simple, we mean simple for events already described from definite events by use of these operations.

A pressing problem now is to consider what kinds of events, described originally in other terms, can be described in these terms; and so eventually what kind of behavior can be explained on the basis of nerve net control.

This is a problem one would naturally investigate in detail next. We have not done so thus far, since this report is

intended only to reduce to writing the author's thinking on the subject during August 1951, and not to try to carry the investigation further, except for the minimum amount of filling in details which was unavoidable in the process of writing.

However, it is very plausible that the notation for regular events in terms of definite events combined by the three operations will prove handy in describing events. The simple examples given at the end of Sect. 7.1, and some others slightly more complicated, encourage this hope.

On the other hand, given a description of an event in terms of definite events and the three operations, it will in some cases be difficult to see what the event consists of; we know of cases in which a very complicated description is actually equivalent to a much simpler one. (This, in fact, is usually the case for descriptions provided by the method of proof of Theorem 6 in Part II.)

So there are problems of translatability in both directions between the notations for regular events and other notations for events or descriptions of events in ordinary language. These problems have so far been touched only superficially, and are crucial for determining how far the present results carry us toward practical general techniques for construction of nets for given purposes.

These questions are related to questions about transformations between different expressions for the same event in terms of our operations. Can we obtain any normal forms, i.e., simplest forms or convenient standard forms, for descriptions

of regular events, to which given forms are equivalent in the sense of Sect. 7.3? Is there any decision procedure for the equivalence of two expressions for events (in the technical sense of modern logic)? These questions are closely related to questions raised in Sect. 7.3.

Similar questions apply to identity in the sense of Sect. 7.3; but equivalence is the important relation for the applications of the theory.

These questions are partly algebraic in character. Some questions are also raised in Part II and Appendix 2.

Success in reducing, to terms of definite events and $E \vee F$, EF, and E*F, events as expressed in ordinary language or as they arise in explaining organic behavior or creating robots for prescribed purposes would, of course, give a justification for our selection of the operations.

Our actual reason for selecting them is that (as was mentioned in Sect. 7.1) a converse of Theorem 4 will be proved in Part II.

Thus, every event which can be represented must be expressible in terms of $E \vee F$, EF, and E*F, starting from definite events.

In particular, we have thus demonstrated that McCulloch-Pitts neurons can govern any kind of behavior which any other kind of digital automaton at all can govern. This, of course, includes a number of special results which they obtained for alternative kinds of nerve nets, but is more general.

Having been first led to the three operations in connection with the converse of Theorem 4 (i.e., Theorem 6 in Part II), it was natural to see whether the present theorem would hold.

But, of course, the fact that our three operations are completely general (by Theorem 6) does not settle the question whether they will prove to be a convenient and practical way to deal with events. Possibly some other selection will prove to be more convenient. Or, we may add other operations and express these in turn in terms of our three.

7.6 Conjunction and negation: We did not include the operations $\&$ (and) and $\overline{\ }$ (not) in our definition of regular events, because in the converse theorem (Theorem 6) we do not need to.

In this section we will show that net constructions can be managed so that the two operations can be included. However, we will only treat them when applied to events already represented by nets, and we will not thereafter use EF and E*F.

From the converse theorem it will follow that any events we thus express using also $\&$ and $\overline{\ }$ must be expressible without them. But the definitions obtained in this way are very complicated, and simple definitions do not appear to be immediately forthcoming. (But we have not examined the problem thoroughly.)

We are not attempting to use $\&$ and $\overline{\ }$ inside EF and E*F (except in the original constructions of definite events as the units) since we have not set up a representation of these operations in terms of classes of definite events or of $\underline{k} \times \underline{\ell}$

tables. It does not seem to be immediate what is the best way to do this.

Theorem 5: Each event constructible from regular events by use of the operations $\&$, $\vee$, and $^-$ of the propositional calculus is representable with lag 2; i.e., a nerve net and a neuron can be found, together with an initial state of the net, so that the neuron fires at time p+2, if and only if the event has taken place ending at time p.

Proof: Say the event is constructed by the operations of the propositional calculus from certain expressions for regular events. Consider any one of the latter. Wherever a part occurs in it of the form E*F, replace this by $F \vee E(E*F)$ using (8). After this, apply (4) and (5) whenever possible. Using also (2) if necessary, we are thus led to an expression for the original event by operations of the propositional calculus in terms of regular parts of the form $E_i(\ldots E_n)$ where E_i is definite; for this purpose we take the $\vee$'s which have been brought outermost in the expressions for the regular events as part of the construction in terms of the operations of the propositional calculus. Say there are m such parts; call their first factors $E_i^{(1)}$ ($1 = 1,\ldots,m$), and the whole expressions $E_i^{(1)}(\ldots E_{n_{(1)}}^{(1)})$. Let $E_i^{(1)\prime\prime}$ be $E_i^{(1)\prime}$ or $E_i^{(1)}$ according as $n_{(1)}$ is > 1 or $= 1$, where $'$ has the meaning given it in the proof of Theorem 4. Now we can take exactly the same combination by operations of the propositional calculus of $E_i^{(1)\prime\prime}, \ldots, E_i^{(m)\prime\prime}$ that the given event is of $E_i^{(1)}(\ldots E_{n_{(1)}}^{(1)}), \ldots, E_i^{(m)}(\ldots E_{n_{(m)}}^{(m)})$. This can

be treated as a definite event of length equal to the greatest length of any of its components, and a net can be constructed for it by Theorem 1, with input neurons $N_1,\ldots,N_k$ and for each i for which $n_{(i)} > 1$ a neuron N_{k+i} required to fire at time $p-\gamma_{(i)}+2$ for the event to occur. Feeding the outputs from the nets for $E_2^{(i)} \ldots E_{n_{(i)}}^{(i)}$ appropriately into this, instead of as before into respective nets for $E_1^{(i)}$, we get a net for the event in question.

PART II — FINITE AUTOMATA:

8. The Concept of a Finite Automaton:

8.1 Cells: Time shall consist (as in Sect. 3 ff.) of a succession of discrete moments numbered as integers. We shall mainly be concerned with the case of only positive integers, as in Sect. 6.2 ff, but will consider the case of all the integers in Appendix 1.

We shall consider automata constructed of a finite number of parts, each being capable of a finite number ≥ 2 of states at any given moment. Call these parts cells.

We shall distinguish two kinds of cells, input cells and inner cells. Say there are k input cells and m inner cells.

An input cell admits 2 states 0 and 1 (or "quiet" and "firing"), which we consider to be determined by the environment.

This restriction to 2 states for input cells is to make the notion of an input to the automaton coincide with the notion of input to a nerve net as formulated in Sect. 4. But the

present theory would work equally well with more than 2 states. Nothing is gained thereby, however, as p cells each admitting 2 states could be used to replace one cell admitting any number q ($2 \leq q \leq 2^p$) states 0, 1, ... , $q-1$, where if $q < 2^p$ we could consider only inputs in which states q, ... , 2^p-1 do not occur or identify all of these states with the state $q-1$ in all the operations of the automaton.

The number of states of an inner cell is not restricted to 2, and different inner cells may have different numbers of states.

Say the input cells are $N_1,\ldots,N_k$ ($k \geq 0$); and the inner cells are $M_1,\ldots,M_m$ ($m \geq 1$), with respective numbers of states $s_1,\ldots,s_m$ (each ≥ 2).

The state of each inner cell at any time t is determined by the states of all the cells at time $t-1$. Of course, it may happen that we do not need to know the states of all the cells at time $t-1$ to infer the state of a given inner cell at time t. Our formulation merely leaves it unspecified what kind of a law of determination we use, except to say that nothing else than the states of the cells at $t-1$ can matter.

If time is given as beginning with $t = 1$, the state of the inner cells at that time is to be specified.

A particular example of a finite automaton is a McCulloch-Pitts nerve net. Here all the cells have just 2 states, and the principles stated in Sect. 3, together with the arrangement of axons and the kind of endbulbs on each case, give the law

determining the state of each inner neuron at time t from the states of all the neurons (or, in fact, from only those having endbulbs synapsing on the given one) at time $t-1$.

Another example is obtained by considering neurons with "alterable synapses" or "alterable endbulbs" of the following kind. Each neuron may have besides the usual endbulbs also exeitatory ones which are not effective unless at some previous time the neuron having the endbulb and the neuron to which the endbulb is adjacent were simultaneously fired. If a neuron has r such alterable endbulbs, it is capable of 2^{r+1} states, according as it is quiet or firing and according to which of the r alterable endbulbs have thus far been made effective.

Many other possibilities suggest themselves.

8.2 State: With input cells $N_1,\ldots,N_k$ and inner cells $M_1,\ldots,M_m$ with respective numbers of states $s_1,\ldots,s_m$, there are possible at a given moment exactly $2^k \cdot s_1 \cdot \ldots \cdot s_m$ states of the entire automaton. We can consider each as a combination of an external state, of which there are 2^k possible, and an internal state of which there are $s_1 \cdot \ldots \cdot s_m$ possible.

The law by which the states of the inner cells at time t are determined by the states of all the cells at time $t-1$ can be given by specifying to each of the complete states at time $t-1$ which one of the inner states at time t shall succeed it.

Now, indeed, there is no reason for our general theory why we cannot consider the entire aggregate of internal cells

as replaced by a single one admitting $s_1 \cdot \ldots \cdot s_m$ states. This normalization of our concept of a finite automaton is always possible, though we did not start out with it, because we were interested in making clear the application to such automata as a McCulloch-Pitts nerve net, where the cells are given certain simple properties and are connected in a certain way.

We could also restrict ourselves to one input cell, by scheduling the inputs on the k original input cells to come in successively in some order on the new one, which would alter the time scale so that k moments of the new time scale correspond to 1 of the original. Events referring to the new time scale could then be interpreted in terms of the original. However, we do not find any advantage in this reduction to one input neuron; so we do not use it.

We will now assume that time starts with $t = 1$. Say we call the states $a_1,\ldots,a_r$ where $r = 2^k \cdot s_1 \cdot \ldots \cdot s_m$ and the internal states $b_1,\ldots,b_q$ where $q = s_1 \cdot \ldots \cdot s_m$. We specify that the internal state at time $t = 1$ be b_1.

Under this assumption, the state at time $t = p$ is a function of the input over time $1,\ldots,p$ (including the value of p, or only this when $k = 0$). (Had we not specified the initial state as b_1 the state at time p would be a function of the initial state also.)

So each of the states $s_1,\ldots,a_r$ corresponds to (or represents) an *event*, which occurs ending at time p, if and only if the input over the time $1,\ldots,p$ is one which results in that one of $a_1,\ldots,a_r$ being the state at time p. Thus, the automaton

can know about the past (inclusive of the present) only that it falls into one of r mutually exclusive classes (possibly some of them empty).

Similar remarks apply to representations of the past by an internal state assumed at time $p+1$, or by a property of the state at time p, or of the internal state at time $p+1$. For to say the internal state at time $p+1$ is b_i means that the complete state at time p was one of certain ones, i.e., those which are succeeded by b_i under the law determining internal state. So then the past falls into a class of possible pasts constituting the set sum of the classes represented by those complete states at time p, or in logical terms the disjunction; similarly, for properties of the state (similarly also, e.g., a property of the internal state at time $p+s$ for $s > 1$, whenever this property does not depend on the input over time $p+1,\ldots,p+s-1$).

What sorts of events can be represented? The question is answered by the following theorem, referring, of course, to automata started in state b_1. Had we not specified the initial state, we would merely add (or disjoin) the classes corresponding thus to the q internal states, each in turn as initial state.

Had we not specified past time to be finite, the state at a given time p would not necessarily be determined by the input. The facts in this case (already mentioned in Sect. 6.1 for McCulloch-Pitts nerve nets) are dealt with in Appendix 1.

As the concept of input is the same as in Part I, we can use the notion of "regular event" which was introduced in Sect. 7.

9. Regularity of Representable Events:

Theorem 6: For any finite automaton (in particular, for a McCulloch-Pitts nerve net) started at a given time $t = 1$ with internal state b_i at that time, the event represented by a given state existing at time p is regular; i.e., the automaton assumes that state at time p, if and only if a certain regular event occurs ending at time p.

Proof: Since the initial internal state is specified, there are 2^k initial states (the results of combining the given initial internal state with each of the 2^k possible external states at time $t = 1$) from which the automaton could start at time $t = 1$ to reach the state in question at time $t = p$.

So if we can show that the automaton can start from a given state at time 1 and reach a given state at time p, if and only if a certain regular event occurs ending at time p, then the theorem will follow by taking the disjunction of 2^k respective regular events, which (by Sect. 7.1) is itself a regular event.

Given any state a at time $t-1$ ($t-1 \geq 1$), exactly 2^k states are possible at time t, since the internal part of the state at time t is determined by a, and the external part can happen in 2^k ways.

So we have a one-many relationship between states. Now invert this relation and consider for any state a at time t what states at time $t-1$ are compatible with it (there may be none, one, or more than one); say a is in relation R to each of these.

The next part of our analysis will apply to any binary relation R defined on a given set of r objects $a_1,\ldots,a_r$ (called "states"), whether or not it arises in the manner just described.

Consider any two a and $\bar{a}$ of the states, not necessarily distinct. We will seek a characterization of the strings of states $d_\chi d_{\chi-1}\ldots d_1$ for which d_χ is a, d_1 is $\bar{a}$, and for each i ($i = 1,\ldots,\chi-1$) d_{i+1} is in the relation R to d_i; call these strings which connect a to $\bar{a}$.

Let A be a class of such strings. We call A regular, if A can be described by an expression built out of the following operations (chosen in analogy to the definition of regular events in Sect. 7.1.)

The empty set and the unit set consisting of just a_i for any i are regular. If A and B are regular, so is their sum which we write $A \vee B$. If A and B are regular, so is the set, written AB, of strings obtained by writing a string belonging to A just left of a string belonging to B. If A and B are regular, so is A*B which abbreviates $\overbrace{A \cdots AB}^{n \text{ factors}}$ ($n \geq 0$), i.e., the sum of these classes for all $n \geq 0$.

Now we prove the lemma by induction on r that the strings $d_\chi\ldots d_1$ connecting a to $\bar{a}$ form a regular class.

Basis: $r = 1$. Then, of course, $\bar{a}$ is a. If $\overline{a \; R \; a}$ (i.e., if not a R a), the class of the connecting strings is simply the unit set consisting of a (as string of length 1), which is regular. If a R a, then the class is $\{a, aa, aaa, \ldots\}$, which is regular, since it can be written A*A where $A = \{a\}$.

Induction step: $\underline{r} > 1$.

Case 1: $a = \bar{a}$. In this case any string leading from a to $\bar{a}$ is of the form

$$a \xrightarrow[\text{no a's}]{} a \xrightarrow[\text{no a's}]{} \cdots \quad a \xrightarrow[\text{no a's}]{} a,$$

Figure 28

where each arrow is either empty (this being possible only if a $\underline{R}$ a), or stands for a string without a in it.

Let $e_1,\ldots,e_{\underline{g}}$ be the states e such that a $\underline{R}$ e, but $e \neq a$, and $f_1,\ldots,f_{\underline{h}}$ the states f such that f $\underline{R}$ a but $f \neq a$. Now any string of the kind represented by the arrow (when the arrow does not represent the absence of a string) must start with one of $e_1,\ldots,e_{\underline{g}}$ and end with one of $f_1,\ldots,f_{\underline{h}}$. For each pair $e_{\underline{i}}f_{\underline{j}}$, by the hypothesis of the induction, the class of the strings leading from $e_{\underline{i}}$ to $f_{\underline{j}}$ (without a in it) is regular. Say $B_1,\ldots,B_{\underline{gh}}$ are these regular classes; let A be $\{a\}$. Now if a $\underline{R}$ a and the B's are not all empty (Subcase i), the class of possible strings a$\longrightarrow$for Fig. 28 is $A \vee A(B_1 \vee \ldots \vee B_{\underline{gh}})$; if a $\underline{R}$ a but all B's are empty (Subcase ii), it is A; if $\overline{a \ \underline{R} \ a}$ and the B's are not all empty (Subcase iii), it is $A(B_1 \vee \ldots \vee B_{\underline{gh}})$; and if a $\underline{R}$ a and the B's are all empty (Subcase iv), it is empty. In the first three subcases, let C denote the class mentioned, which is non-empty and regular. Now in these subcases, the class of strings leading from a to a (as in Fig. 28) is C*A, while in the fourth subcase it is simply A.

Case 2: $a \neq \bar{a}$. Now we have instead of Fig. 28 the following:

$$a \xrightarrow[\text{no } a\text{'s}]{} \quad a \xrightarrow[\text{no } a\text{'s}]{} \quad \cdots \quad a \xrightarrow[\text{no } a\text{'s}]{} \quad a \xrightarrow[\text{no } a\text{'s or } \bar{a}\text{'s}]{}$$

$$\bar{a} \xrightarrow[\text{no } a\text{'s or } \bar{a}\text{'s}]{} \quad \cdots \quad \bar{a} \xrightarrow[\text{no } a\text{'s or } \bar{a}\text{'s}]{} \quad \bar{a}$$

Figure 29

The treatment is similar. For example, in the case the classes of strings represented by "$a \xrightarrow[\text{no } a\text{'s}]{}$", by "$a \xrightarrow[\text{no } a\text{'s or } \bar{a}\text{'s}]{}$" in the middle, and by "$\bar{a} \xrightarrow[\text{no } a\text{'s or } \bar{a}\text{'s}]{}$" at the right, are none of them empty, call them C, D, and E, respectively, the class of strings for Fig. 29 is $C^*ADE^*\bar{A}$, where $\bar{A} = \left\{\bar{a}\right\}$.

So the lemma is proved. Now we return to the point where we were in the proof of the theorem. We wish to show that for given state a at time p and each of 2^k possible states $\bar{a}$ at time 1, that a holds at $t = p$ and $\bar{a}$ at $t = 1$, if and only if a certain proper event (different in general for each of the 2^k $\bar{a}$'s) occupies the time 1 to p.

Now by the lemma, the strings of states which can lead from a to $\bar{a}$ form a regular class. If that class is empty, then the event is the improper one, which is regular. If that class is not empty, consider the expression for the class as a regular class of strings. We build a corresponding expression for the event as a regular event by translating each unit class $A_i = \left\{a_i\right\}$ (for each state a_i) into the definite event E_i

of length 1 which occurs at time $\underline{p}$ exactly if the input at time $\underline{p}$ is the external part of the state $a_{\underline{i}}$. (If $\underline{k} = 0$, $E_{\underline{i}}$ is the event I of length 1 which always occurs, having no other features. However, it may be initial 1, that is, $E^{o}_{\underline{i}}$. The only other event of length 1 in this case is the improper one $\bar{I}$.) Having done this, then the operations $E \vee F$, EF, and $E^{*}F$ for building regular events parallel those $A \vee B$, AB, and $A^{*}B$ for building regular classes of strings. The earliest units (Sect. 7.3) in the expressions obtained should be marked as being initial.

This proves the theorem. No attempt has been made to consider, in this proof, how simply the event represented by the state at time $\underline{p}$ can be constructed as a regular event. We have worked out some simple illustrations in which very complicated expressions stand for regular events capable of simple ones. The expressions obtained have entirely initial events as earliest units, and are built of units of length 1. It is clear that in most examples great simplifications can be obtained by use of equivalences (Sect. 7.3); but no study has yet been made of the possibilities for proceeding systematically with such simplifications, or of rearranging the proof of Theorem 6 to come out directly with simpler expressions when possible.

The study of the structure of a set of objects $a_1,\ldots,a_{\underline{r}}$ under a relation $\underline{R}$, which is at the heart of the above proof, might profitably draw on some algebraic theory, since it is possible (though whether profitable or not we do not know) to see the situation as a generalization of permutation groups.

Corollary: The event ending at time p represented by each of the following is likewise regular: (a) a certain internal state at time p+1. (b) a property of the state at time p. (c) a property of the internal state at time p+1. (d) a certain internal state at time p+s for a given s > 1, when this does not depend on the input over p+1 to p+s-1. (e) a property of the internal state at time p+s for a given s > 1, when this does not depend on the input over p+1 to p+s-1.

Proof: As remarked at the end of Sect. 8, each of these is equivalent to one of certain states existing at time p; so the event represented is a disjunction of the regular events given by the theorem for the latter, and hence is regular.

This corollary brings our result now into correspondence (as converse) of the result in Sect. 7. There we represented an event by firing a certain neuron at time p+2. This is a property of the internal state at time p+2, since it means the internal state then is one of 2^{m-1} different ones (according to the states of the other m-1 inner neurons).

Incidentally, the representations in Sect. 7 by firing a certain neuron at time p+2 are equivalent to representations by a property of the state at time p, namely by the property which those states at time p share which will lead to the firing of this neuron at time p+2.

The event which is represented by a state a is the solution in the sense in which McCulloch-Pitts speak of "solution" for the case of nerve nets, except that we give the solution

for a given internal initial state. A solution without presupposing an initial internal state would then be obtained as a disjunction of the solutions for us for each of the $\underline{s}_1 \cdot \ldots \cdot \underline{s}_m$ (for nerve nets, $2^{\underline{m}}$) internal states.

Appendix 3 contains an example of an event which cannot be represented in a finite automaton.

It is, of course, essential for our arguments here that the number of cells or parts (under our first definition of a finite automaton) and the number of states for each, be finite, so that the number of complete states is fixed in advance. A machine of Turing (1937) which is supplied with an unlimited amount of tape, is not a finite automaton in our present sense, since, although in its operation only a finite number of squares of tape are printed upon at any time, there is no preassigned bound to this number.

The Turing machine could be thought of as a finite automaton, which is also able to store information in the environment and reach for it later, so that in certain cases the inputs are identified with inputs at earlier times or with states of certain inner cells at earlier times, and thus the present input is not entirely independent of the past. Whether this comparison may lead to any useful insights into Turing machines, or reciprocally into finite automata, remains undetermined.

APPENDIX 1: DEFINITENESS OF EVENTS REPRESENTABLE IN A FINITE AUTOMATON WITH AN INFINITE PAST:

Theorem 7: Every event E ending at time p representable by a certain state existing at time p (or by one of the other methods listed in Corollary Theorem 6) in a finite automaton with an infinite past is definite.

The result was cited in Sect. 6.1. The notion of automaton to which we refer is given in Sect. 8.

Proof: With k input cells, the complete past is generated by choosing between a finite number of 2^k possible inputs at time $t = p$, then between a finite number 2^k at time $t = p-1$, etc., ad infinitum.

By a theorem given by Brouwer (1924) and also by Konig (1927), if for each infinite past (i.e., for any such choice sequence) it is determined at some finite stage whether the event occurs (ending at time p) or not, then there must be a number $N \geq 0$ such that, whether the event occurs or not is known for all pasts (i.e., all choice sequences) from only that part of the past occupying the time $p-N$ to p. In this case the event would be definite of length N+1. (Brouwer's proof of the theorem is intended for readers acquainted with the intuitionistic set theory, and the main effort in his proof is to demonstrate the theorem intuitionistically.)

Now we show that indefinite events are not representable. Contraposing the mentioned theorem, we conclude that for an indefinite event, there is some particular infinite past such that for every u it is not known from the part from $p-u$ to p of the past whether the event occurs or not.

Case 1: The event E does not occur for this particular past. Then for every u there is a past coinciding with the given one over the time $p-u$ to p and diverging from it prior to $p-u$, for which the event occurs.

Now suppose the representation of the event is by a property of the state at time p; say the notation is arranged so that the states which have this property are $a_1,\ldots,a_{r_1}$ and the states which do not are $a_{r_1+1},\ldots,a_r$.

Now consider the set S of all the sequences of states $d_0d_1d_2\ldots$ compatible with the present state being one of $a_{r_1+1},\ldots,a_r$; i.e., d_0 is one of $a_{r_1+1},\ldots,a_r$, and each d_i has as its internal part that which is determined by d_{i+1}. There are $r-r_1$ choices for d_0, at most r for d_1, at most r for d_2, etc.

But for any u there is a past coinciding with the given one over the time $p-u$ to p, and diverging from it before that, along which the event does not occur. Along this past, any sequence $d_0d_1d_2\ldots$ must belong to S and must in its first $u+1$ choices $d_0d_1\ldots d_u$ be compatible with the given past (as selected above); i.e., the external parts of d_0, d_1, $\ldots,d_u$ must be the inputs over the last $u+1$ moments of that past.

Now by Brouwer's theorem (contraposed) there must be an infinite sequence $d_0d_1d_2\ldots$ in the set S which is compatible with the entire given past (along which E occurs, but from every finite segment of which a past diverges along which E does not occur).

But d_0 is one of the states $a_{\underline{r}_1+1},\ldots,a_{\underline{r}}$, contrary to our assumption that E is represented by the state at $\underline{t} = \underline{p}$ being one of $a_1,\ldots,a_{\underline{r}_1}$. Thus, E cannot be so represented.

If we had assumed simply that whenever E occurs, the present state must be one of $a_1,\ldots,a_{\underline{r}_1}$, the above considerations show that there must also be examples in which the present state will be one of $a_1,\ldots,a_{\underline{r}_1}$ without E having occurred.

Case 2: The event E does not occur for this particular past. The reasoning already applied gives the absurdity of $\bar{E}$ being represented, hence of E itself being represented, by a property of the state at $\underline{t} = \underline{p}$.

APPENDIX 2: PRIMITIVE RECURSIVENESS OF REGULAR EVENTS:

Theorem 8: Every regular event is primitive recursive.

The terminology in the theorem is from the theory of recursive functions and predicates as developed in the last 25 years. A book by Péter summarizes the theory, also a book by the author which it is hoped will soon be in print.

The formulas given below "place" the regular events in relation to number-theoretic predicates studied in the theory of recursive functions. Although we have not pursued the matter further than to get one way of expressing regular events recursively, possibly useful characterizations may be obtainable in this direction.

We already know from Sects. 5, 6.3, and 7.1 that number-theoretic formulas can be constructed to stand for definite events. The symbolism required can be seen by inspection of the examples given. Terms $p-1$, $p-2$, ... are used only when they are greater than 0, as is insured by adding $p \geq \chi$ (Sect. 6.3) or now sometimes $p = \chi$ (Sect. 7.1) to the expressions as given in Sect. 5.

The range of the variables in the theory of recursive functions is customarily 0, 1, 2, ... rather than 1, 2, 3, ... To avoid having to reconstruct the notations in that theory for the present application, let us in this appendix suppose the time scale for events starts with $t = 0$ instead of $t = 1$. Slight changes are then required in the formulas for definite events. Incidentally, now $p-1$, $p-2$, ... are used in the sense of $p \dot{-} 1$, $p \dot{-} 2$ in the theory of recursive functions.

We shall simultaneously build expressions for regular events and for the lengths of the definite events of the class of definite events which we use in characterizing the regular events. More precisely, we describe to each regular event E expressions $\underline{A}(\underline{p})$, $\underline{A}(\underline{p},\underline{n})$, $\epsilon(\underline{p})$, and $\mu(\underline{n})$ such that

$$\left\{\underline{E} \text{ has occurred ending at time } \underline{p}\right\} \equiv (\underline{En})_{\underline{n}<\epsilon(\underline{p})}\underline{A}(\underline{p},\underline{n}) \equiv \underline{A}(\underline{p}),$$

where for each $\underline{n}<\epsilon(\underline{p})$, $\underline{A}(\underline{p},\underline{n})$ describes a definite event of length $\mu(\underline{n})$. Here $\epsilon(\underline{p})$ is the number (≥ 1) of definite events, in the occurrence of one of which the regular event consists. (Our $\epsilon(\underline{p})$ is not necessarily the least number of such definite events, but is the number we use in our construction. Also, the $\epsilon(\underline{p})$ regular events need not all be different.)

For a definite event used as a unit in the construction of a regular event, $\epsilon(\underline{p}) = 1$ (so $\underline{n}$ has only one value 0); and $\underline{A}(\underline{p})$ is the expression already mentioned for the event; $\underline{A}(\underline{p},\underline{n})$ can be simply $\underline{A}(\underline{p})$, or $\underline{A}(\underline{p}) \,\&\, \underline{n} = 0$ if we wish $\underline{n}$ to appear explicitly in this case; while $\mu(\underline{o}) = \underline{\ell}$ where $\underline{\ell}$ is the length of the definite event under consideration.

For a regular event of the form $F \vee G$, say that $\underline{B}(\underline{p},\underline{n})$, $\zeta(\underline{p})$ and $\nu(\underline{n})$ are the "$\underline{A}(\underline{p},\underline{n})$", "$\epsilon(\underline{p})$" and "$\mu(\underline{n})$" for F, and $\underline{C}(\underline{p},\underline{n})$, $\eta(\underline{p})$ and $\xi(\underline{n})$ are those for G.

Now take

$$\underline{A}(\underline{p}) \equiv (\underline{En})_{\underline{n}<\epsilon(\underline{p})}\, \underline{A}(\underline{p},\underline{n})$$

$$\equiv (\underline{En})_{\underline{n}<\epsilon(\underline{p})} \left[\left\{\underline{n}<\zeta(\underline{p}) \,\&\, \underline{B}(\underline{p},\underline{n})\right\} \vee \left\{\underline{n}\geq\zeta(\underline{p}) \,\&\, \underline{C}(\underline{p},\underline{n}\dot{-}\zeta(\underline{p}))\right\}\right]$$

(the scope of the prefix $(\underline{En})_{\underline{n}<\epsilon(\underline{p})}$ in the last being $\underline{A}(\underline{p},\underline{n})$),

where

$$\epsilon(\underline{p}) = \zeta(\underline{p}) + \eta(\underline{p}),\ \mu(\underline{n}) = \begin{cases} \nu(\underline{n}) \text{ if } \underline{n} < \zeta(\underline{p}), \\ \xi(\underline{n} \dot{-} \zeta(\underline{p})) \text{ if } \underline{n} \geq \zeta(\underline{p}). \end{cases}$$

Let $[\underline{a}/\underline{b}]$ = the quotient, and $\rho(\underline{a},\underline{b})$ = the remainder, when an integer $\underline{a}$ is divided by a positive integer $\underline{b}$. Note that as $\underline{a}$ ranges from 0 to $\underline{kb}-1$ $(b \neq 0)$, the pair of quantities $[\underline{a}/\underline{b}]$, $\rho(\underline{a},\underline{b})$ range over all pairs of numbers $\underline{x},\underline{y}$ with $\underline{x} < \underline{k}$, $\underline{y} < \underline{b}$.

Now for FG, given expressions for F and for G as before by the hypothesis of the induction, we have

$$\underline{A}(\underline{p}) \equiv (\underline{En})_{\underline{n} < \epsilon(\underline{p})}\ \underline{A}(\underline{p},\underline{n})$$

$$\equiv (\underline{En})_{\underline{n} < \epsilon(\underline{p})} \left\{ \underline{B}\left(\underline{p}, [\underline{n}/\eta(\underline{p})]\right) \,\&\, \underline{C}\left(\underline{p} \dot{-} \nu([\underline{n}/\eta(\underline{p})]),\ \rho(\underline{n}, \eta(\underline{p}))\right) \right\},$$

where $\quad \epsilon(\underline{p}) = \zeta(\underline{p})\ \eta(\underline{p})$

$$\mu(\underline{n}) = \nu([\underline{n}/\eta(\underline{p})]) + \xi(\rho(\underline{n},\eta(\underline{p}))).$$

For the remaining case we use the function $(\underline{a})_{\underline{i}}$ defined thus. First let $\underline{p}_{\underline{i}}$ = {the $\underline{i}$-th prime counting 2 as the 0-th}. (So $\underline{p}_0 = 2$, $\underline{p}_1 = 3$, $\underline{p}_5 = 13$, etc.) Now

$$(\underline{a})_{\underline{i}} = \begin{cases} \text{the highest power of } \underline{p}_{\underline{i}} \text{ which divides } \underline{a}, \text{ if } \underline{a} \neq 0, \\ 0 \text{ if } \underline{a} = 0. \end{cases}$$

Note that $(\underline{a})_{\underline{i}}$ is also 0 if $\underline{a} \neq 0$, but $\underline{p}_{\underline{i}}$ does not divide $\underline{a}$. For example, $28 = 2^2 \cdot 7$; so $(28)_0 = 2$, $(28)_1 = 0$, $(28)_2 = 0$,

$(28)_3 = 1$, and $(28)_{\underline{i}} = 0$ for any $\underline{i} > 3$.

As $\underline{a}$ ranges over all non-negative integers, $(\underline{a})_0,\ldots,(\underline{a})_{\underline{m}-1}$ range over all $\underline{m}$-tuples of natural numbers; and, in fact, as $\underline{a}$ ranges over 0 to $p_0^{b_0} \cdot \ldots \cdot p_{\underline{m}-1}^{b_{\underline{m}-1}}$ or beyond the functions $\max((\underline{a})_0,\underline{b}_0),\ldots,\max((\underline{a})_{\underline{m}-1},\underline{b}_{\underline{m}-1})$ range over all $\underline{m}$-tuples $x_0,\ldots,x_{\underline{m}-1}$ for $\underline{x}_0 \leq \underline{b}_0,\ldots,\underline{x}_{\underline{m}-1} \leq \underline{b}_{\underline{m}-1}$. (The function $(\underline{a})_{\underline{i}}$ could have been used in place of $\left[\underline{a}/\underline{b}\right]$ and $\rho(\underline{a},\underline{b})$ in treating the preceding case.)

Now say E is F*G, where expressions as before are assumed already constructed for F and for G. An occurrence of E is an occurrence of $\overbrace{F \ldots F}^{\underline{u} \text{ factors}}FG$ for some $\underline{u} \geq 0$. But for a given $\underline{p}$, we must have $\underline{u} \leq \underline{p}$, since F and G are each of length ≥ 1. So now we have for E*F the following:

$$\underline{A}(\underline{p}) \equiv (\underline{E}\underline{n})_{\underline{n} < \epsilon(\underline{p})} \underline{A}(\underline{p},\underline{n})$$

$$\equiv (\underline{E}\underline{n})_{\underline{n} < \epsilon(\underline{p})} \Big[(\underline{i})_{1 \leq \underline{i} \leq \max((\underline{n})_0,\ \underline{p})} \underline{B}(\max((\underline{n})_{\underline{i}},\ \zeta(\underline{p}) \dot{-} 1),\ \underline{p} \dot{-} \sum_{\underline{s}=1}^{\underline{i}} \gamma(\max((\underline{n})_{\underline{s}},\ \zeta(\underline{p}) \dot{-})))$$

$$\& \ \underline{C}(\max((\underline{n})_{\max((\underline{n})_0,\underline{p})+1},\eta(\underline{p}) \dot{-} 1),\ \underline{p} \dot{-} \sum_{\underline{s}=1}^{\max((\underline{n})_0,\underline{p})} \gamma(\max((\underline{n})_{\underline{s}},\ \zeta(\underline{p}) \dot{-} 1)))\Big],$$

where

$$\epsilon(\underline{p}) = p_{\underline{p}+1}^{\underline{p}+\zeta(\underline{p})+\eta(\underline{p})}$$

since each $\underline{n}$ which would be wanted is of the form

$$2^{\underline{u}} \cdot \underline{p}_1^{\underline{v}_1} \cdot \ldots \cdot \underline{p}_{\underline{u}}^{\underline{v}_{\underline{u}}} \cdot \underline{p}_{\underline{u}+1}^{\underline{w}} ,$$

where

$$\underline{u} \leq \underline{p},$$

$$v_1, \ldots, \underline{v}_{\underline{u}} < \zeta(p) ,$$

$$\underline{w} < \eta(\underline{p}) ,$$

and

$$\mu(\underline{n}) = \sum_{\underline{s}=1}^{\max((\underline{n})_0, \underline{p})} \nu(\max((\underline{n})_{\underline{s}}, \zeta(\underline{p}) \dot{-} 1)) +$$

$$\xi(\max((\underline{n})_{\max((\underline{n})_0, \underline{p})+1}, \eta(\underline{p}) \dot{-} 1)).$$

It will be seen by readers familiar with recursive function theory that $\{\underline{E}$ occurs ending at time $\underline{p}\}$ is thus primitive recursive in the predicates $\underline{N}_1(\underline{t}), \ldots, \underline{N}_{\underline{k}}(\underline{t})$ giving the inputs over time $\underline{t} = 0, \ldots, \underline{p}$, though, of course, the recursive expressions are complicated. Also, we can express the result by saying $\{\underline{E}$ occurs ending at time $\underline{p}\}$ is primitive recursive in $\underline{p}$ and a number giving in code form the combined input from $\underline{t} = 0$ to $\underline{t} = \underline{p}$.

APPENDIX 3: AN EXAMPLE OF AN EVENT WHICH IS NOT REPRESENTABLE (AND THEREFORE NOT REGULAR), THOUGH IT IS PRIMITIVE RECURSIVE:

Consider the event E referring to one input cell N, described as follows:

N fires at time $t = v^2$ for every v such that $v^2 \leq p$, and only then.

(This is primitive recursive, since it can be expressed thus:

$$(u)_{u\leq p}\left\{\left[(Ev)_{v\leq p}\left[u=v^2\right] \,\&\, N(u)\right] \vee \left[(\overline{Ev})_{v\leq p}\left[u=v^2\right] \,\&\, \overline{N(u)}\right]\right\}.)$$

No nerve net or finite automaton of any other kind can represent this event. For consider an automaton with states $a_1,\ldots,a_r$.

Assume given a representation of the event by a property of the automaton at time p; i.e., we assume that there are states, say $a_1,\ldots,a_{r_1}$ ($r_1 < m$), such that at time p the state is or is not one of these, according as the event has occurred or not.

We shall show that this assumption leads to absurdity.

Consider any number s such that $2s > r_1$.

Say that N is fired at times $t = 1, 4, 9,\ldots,s^2$ and never thereafter.

Then E occurs for $p = 1, 2,\ldots,(s+1)^2 - 1$ and for no greater p.

Consider the states of the automaton at times s^2+1, s^2+2, ... These must all be from the list $a_1,\ldots,a_r$.

However, beginning with time $\underline{s}^2+1$, N is never fired; so the external state is constant. Thus, each state for all time thereafter will be determined by the immediately preceding state. Hence, since there is only a finite number of possible states $a_1,\ldots,a_{\underline{r}}$, the sequence of the states $d_1,d_2,d_3,\ldots$ beginning with that at time $\underline{s}^2+1$ is ultimately periodic. For after $\underline{r}$ states at most, a state must be taken for the second time, and thereafter the states since the first occurrence of that one must repeat themselves cyclically.

However, during the time $\underline{s}^2+1,\ldots,\underline{s}^2+2\underline{s}$, the state must be one of $a_1,\ldots,a_{\underline{r}_1}$, since the event occurs for these values of $\underline{p}$; and hence, since $2\underline{s} > \underline{r}_1$, the period must already have become established (i.e., the first repetition in $d_1,d_2,d_3,\ldots$ must already have occurred) by the time $\underline{s}^2+2\underline{s}$. Hence the state at time $\underline{s}^2+2\underline{s}+1$ $(= (\underline{s}+1)^2)$ is one of $a_1,\ldots,a_{\underline{r}_1}$, contrary to the fact that the event has not occurred ending at time $\underline{p} = (\underline{s}+1)^2$.

It is not suggested that the event in question would be of any biological significance. But the example is given to show the mathematical limitations to what events can be represented. Of course, by Appendix 2 we already knew that events not primitive recursive are not representable; but the present example is much simpler.

Without either appendix, one would not expect events whose verification involves the completion of an infinite process (these being non-recursive) to be representable. The

present example does not involve the completion of an infinite process; but it does involve the completion of a finite process, which as p varies is unbounded, and this likewise transcends the capabilities of a fixed finite automaton.

bjc

BIBLIOGRAPHY

Brouwer, L. E. J. Beweis, dass jede volle Funktion gleichmässig stetig ist. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Amsterdam, Vol. 27 (1924), pp. 189–193. (Also in another version in Über Definitionsbereiche von Funktionen, Mathematische Annalen, Vol. 93 (1927), pp. 60–75.)

Culbertson, James T. Consciousness and Behavior. William C. Brown Company, Dubuque, Iowa, 1950.

Hilbert, David, and Ackermann, Wilhelm. Grundzüge der Theoretischen Logik. First Ed., Berlin (Springer) 1928, viii + 120 pp. Second Ed., same 1938; reprinted by Dover 1944; translated and printed in English recently. Third Ed. in German recently printed in Germany.

Householder, A. S., and Landahl, H. D. Mathematical Biophysics of the Central Nervous System. Mathematical Biophysics Monograph Series, No. 1, Principia Press, Bloomington, Indiana, 1945.

Kleene, S. C. Introduction to Metamathematics. Forthcoming.

König, D. Über eine Schlussweise aus dem Endlichen ins Unendliche. Acta Litt. ac. Scient. Univ. Szeged, Sect. Math. Vol. III/II (1927) pp. 121–130; particularly the appendix, pp. 129–130.

McCulloch, Warren S. "The Brain as a Computing Machine," Electrical Engineering, Vol. 68 (1949), pp. 492–497.

McCulloch, Warren S., and Pitts, Walter. "A Logical Calculus of the Ideas Immanent in Nervous Activity," Bulletin of Mathematical Biophysics, Vol. 5 (1943), pp. 115–133.

Peter, Rozsa. Rekursive Funktionen. Akademiai Kiado, Adademischer Verlag, Budapest, 1951. 206 pp.

Turing, A. M. "On Computable Functions with an Application to the Entscheidungsproblem," Proceedings of the London Mathematical Society, Ser. 2, Vol. 42 (1937), pp. 230–265. "A Correction," Same, Vol. 43 (1937), pp. 544–546.