XVI. NEUROPHYSIOLOGY*

W. S. McCulloch · R. C. Gesteland · W. H. Pitts
E. M. Duchane · J. Y. Lettvin · P. D. Wall

A. THREE OF VON NEUMANN'S BIOLOGICAL QUESTIONS

Neurophysiologists are indebted to John von Neumann for his studies of components and connections in accounting for the steadiness and the flexibility of behavior. In speaking to the American Psychiatric Association (1) he stressed the utility and the inadequacy of known mechanisms for stabilizing nervous activity, namely, (a) the threshold of nonlinear components, (b) the negative feedback of reflexive mechanisms, (c) the internal switching to counteract changes—"ultrastability"—(2), and (d) the redundancy of code and of channel. He suggested that the flexibility might depend upon local shifts of thresholds or incoming signals to components that are more appropriate to computers than any yet invented. His Theory of Games (3) has initiated studies that may disclose several kinds of stability and has indicated where to look for logical stability under common shift of threshold. His "Toward a Probabilistic Logic" (4) states the problem of securing reliable performance from unreliable components, but his solution requires better relays than than he could expect in brains. These, his interests, put the questions we propose to answer. His satisfaction with our mechanisms for realizing existential and universal quantification in nets of relays (5, 6) limits our task to the finite calculus of propositions. Its performance has been facilitated by avoiding the opacity of the familiar symbols of logic and the misleading suggestions of multiplication and addition modulo two of the facile boolean notation for an algebra that is really substitutive (7, 8, 9). Our symbols have proved useful in teaching symbolic logic in psychological and neurological contexts (10). Familiarity with them undoubtedly contributed to the invention of the circuits whose redundancy permits solution of our problems. See Fig. XVI-1.

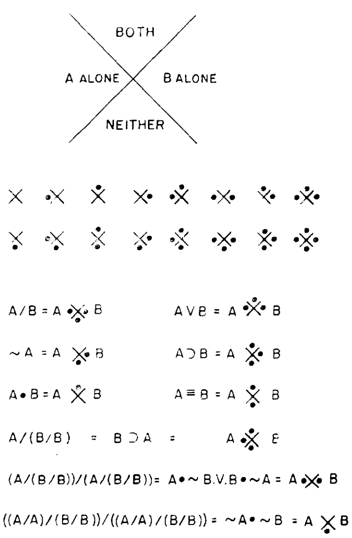

Fig. XVI-1. Symbols.

The finite calculus of propositions can be written at great length by repetitions of a stroke signifying the incompatibility of its two arguments. The traditional five symbols, for 'not', 'both', 'or', 'implies', and 'if and only if', shorten the text

*This work was supported in part by Bell Telephone Laboratories, Incorporated; in part by The Teagle Foundation, Incorporated; and in part by National Science Foundation.

but require conventions and rearrangements in order to avoid ambiguities. Economy requires one symbol for each of the sixteen logical functions of two propositions. The only necessary convention is then one of position or punctuation.

Since the logical probability and the truth value of a propositional function are determined by its truth table, each symbol should picture its table. When the place in the table is given, any jot serves for "true" and a blank for "false." When the four places in the binary table are indicated by a cross (×) it is best to let the place to the left show that the first proposition alone is the case; to the right, the second; above, both; and below, neither. Every function is then pictured by jots for all of those cases in which the function is true, ranging from contradiction, with no jots, to tautology, with four.

Formulas composed of our symbols are transparent when the first proposition is represented by a letter to the left of the symbol and the second to the right. When these spaces are occupied by logical variables the formula is that of a propositional function; when they are occupied by propositions, of a proposition; consequently the formula can occupy the position of an argument in any subsequent formula.

Two distinct propositions, A and B, are independent when the truth value of either does not logically determine the truth value of the other. A formula with only one symbol whose spaces are occupied by two independent propositions can never have the symbol with no jots or four jots. The truth value of any other function is contingent upon the truth value of its arguments. Let us call it "a significant proposition of the first rank."

A formula for a proposition of the second rank is formed by inserting in the spaces of its symbol two significant propositions of the first rank; for example, (A × B) × (A × B). When the two propositions of the first rank are composed of the same pair of propositions in the same order, the resulting function of the second rank can always be equated to one of the first rank; for example, (A × B) × (A × B) = (A × B), by putting jots into the cross by the four rules of reduction:

1) When the central × has a jot at the left, insert a jot in the new × for every jot in the × to the left but not in the × to the right. (A × B) × (A × B) = (A × B).
2) When the central × has a jot at the right, insert a jot in the new × for every jot in the × to the right but not in the × to the left. (A × B) × (A × B) = (A × B).
3) When the central × has a jot above, put a jot in the new × for all jots common to the right and left ×'s. (A × B) × (A × B) = (A × B).
4) When the central × has a jot below, put a jot in the new × for every space empty in both the right and left ×'s. (A × B) × (A × B) = (A × B).

By repetition of the construction we can produce formulas for functions of the third and higher ranks and reduce them step by step to the first rank, thus discovering their truth values.

Since in what follows no other formulas are necessary, the letters A and B will be omitted, and positions, left and right, will replace parentheses.

In formulas for significant functions of the first rank the chance addition or omission of a jot produces an erroneous formula and will cause an error only in that case for which the jot is added or omitted, which is one out of the four logically equiprobable cases. With similar symbols for functions of three arguments, the error will occur in only one of the eight cases, and, in general, for functions of δ arguments, in one of 2^{δ} cases. If p is the probability of the erroneous jot and P the probability of error produced, $P = 2^{-\delta} p$.

In formulas for the second rank there are three symbols. If we relax the requirement of independence of the arguments, A and B, there are then 16^3 possible formulas each of which reduces to a formula of the first rank. Thus the redundancy, R, of these formulas of the second rank is $16^3 \div 16 = 16^2$. For functions of δ arguments, $R = (2^{2^{\delta}})^{\delta}$.

To exploit this redundancy so as to increase the reliability of inferences from unreliable symbols let us realize the formulas in nets of what von Neumann called neurons (3). Each neuron is a relay which on receipt of all-or-none signals either emits an all-or-none signal or else does not emit one which it would otherwise have emitted. Signals approaching a neuron from two sources either do not interact, or those from one source prevent some or all of those from the other source from reaching the recipient neuron. The diagrams of the nets of Fig. XVI-2 are suggested by the anatomy of the brain. They are to be interpreted as follows.

A line terminating upon a neuron shows that it excites it with a value +1 for each termination. A line forming a loop at the top of the neuron inhibits it with a value of excitation of −1 for each loop. A line forming a loop around a line approaching a neuron shows that it prevents excitation or inhibition from reaching the neuron through that line.

Each neuron has on any occasion a threshold, θ, measured in steps of excitation, and it emits a signal when the excitation it receives is equal to or greater than θ. The output of the neuron is thus some function of its input, and which function it is depends upon both its local connections and the threshold of the neuron. These functions can be symbolized by crosses and jots beginning with none and adding one at a time as θ decreases until all four have appeared in the sequence noted in the legend for its diagram in Fig. XVI-2. Because all 24 sequences (of which only 12 left-handed are drawn) are thus realized we can interpret the accidental gain or loss of a jot or jots in an intended symbol as a change in the threshold of an appropriate neuron.

The formula for a function of the second rank is realized by a net of three neurons each of whose threshold is specified; for example, see Fig. XVI-3. The function can be reduced to one of the first rank whose symbol pictures the relation of the output of the net to the input of the net.

When all thresholds shift up or down together, so that each neuron is represented by one more, or one less, jot in its symbol but the reduced function is unaltered, the

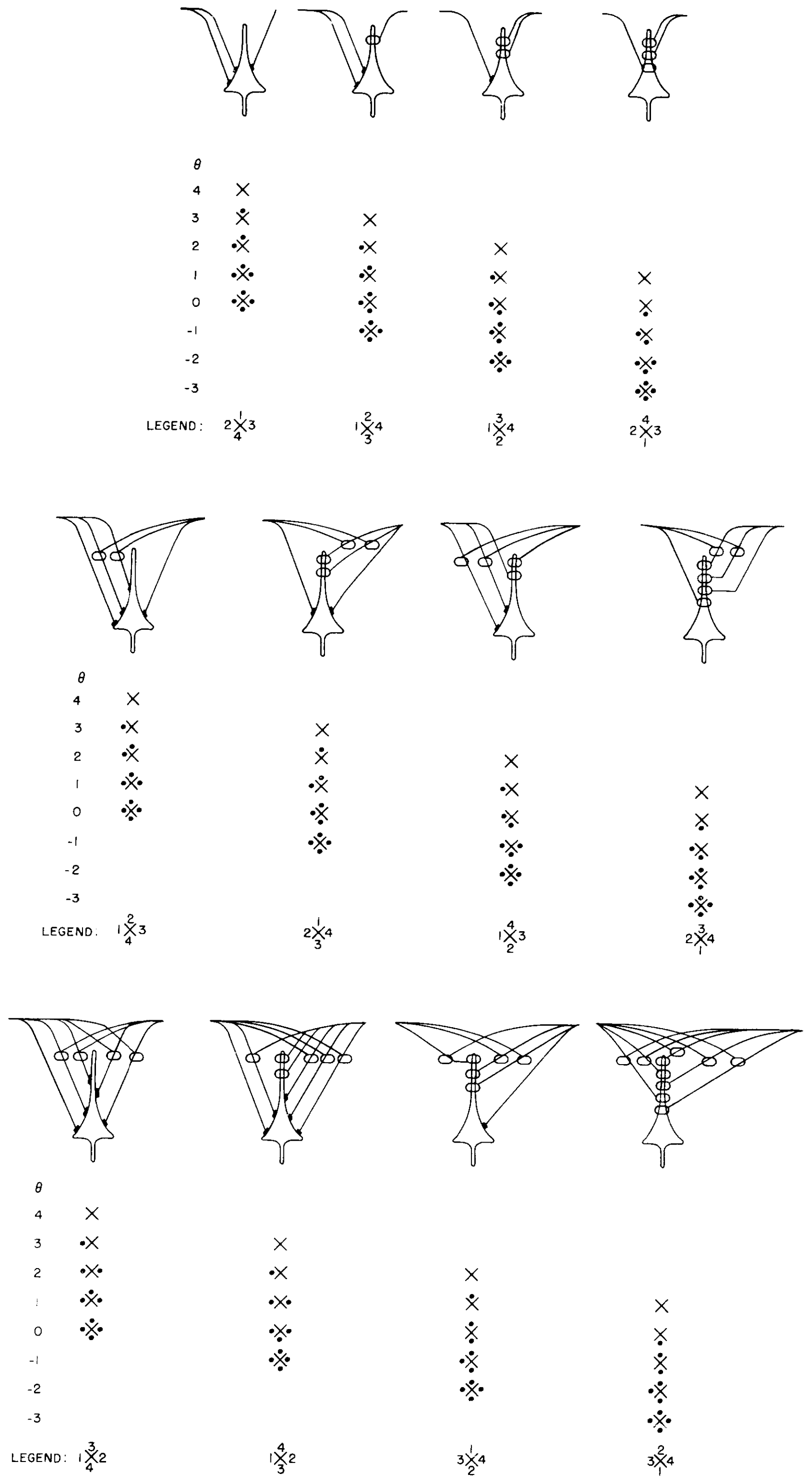

Fig. XVI-2. Diagrams.

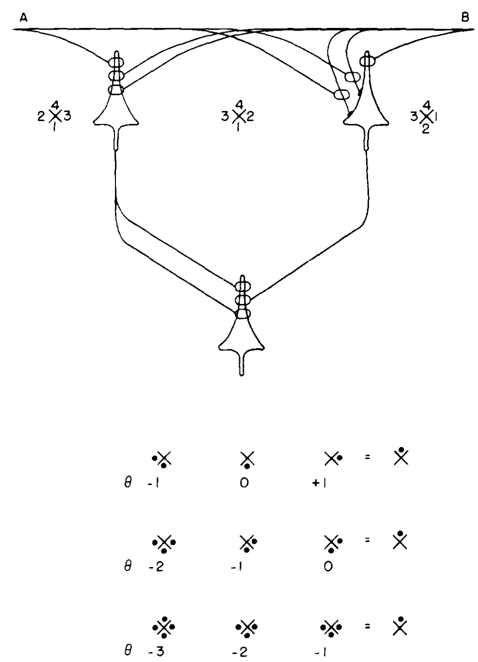

Fig. XVI-3. A logically stable net.

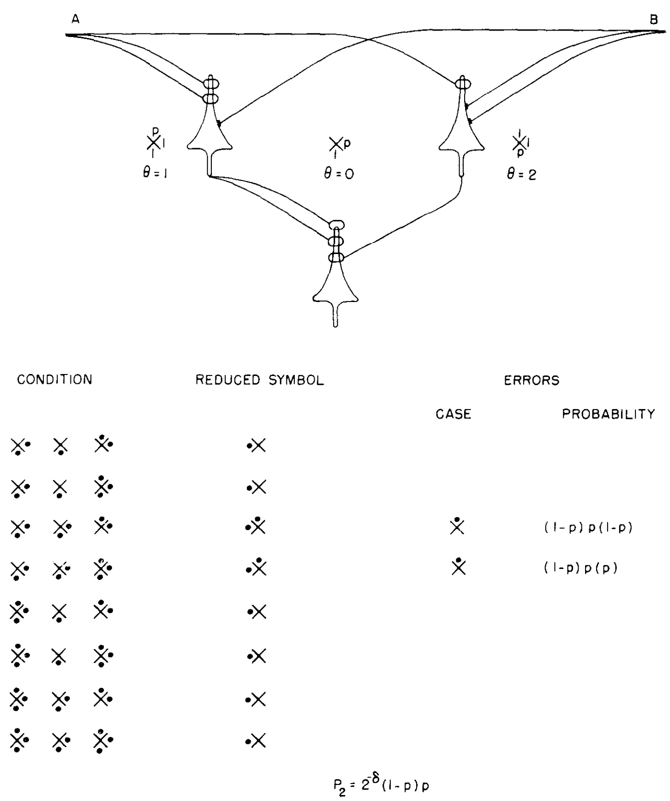

Fig. XVI-4. A best stable net.

net is called "logically stable."

The redundancy of functions of the second rank provides us with many examples of pairs of functions and even triples of functions that reduce to the same function of the first rank and that can be made from one another by common addition or omission of jots in all symbols, and the diagrams of Fig. XVI-2 enable us to realize them all in several ways: For example, there are 32 triples of functions and 64 logically stable nets for every reduced formula with a single jot.

If such nets are embodied in our brains they answer von Neumann's repeated question of how it is possible to think and to speak correctly after taking enough absinthe or alcohol to alter the threshold of every neuron. The limits are clearly convulsion and coma, for no function remains significant or stable with a shift of θ that reduces the output neuron to tautology or to contradiction. The net of Fig. XVI-3 is logically stable over the whole range between these limits. Let the causes and probabilities of such shifts be what they may, those that occur simultaneously throughout these nets create no errors.

Logically stable nets differ greatly from one another in the number of errors they produce when thresholds shift at random in their neurons and the most reliable make some errors; for example, the net of Fig. XVI-4.

To include random shifts, let our symbols be modified by replacing a jot with 1 when the jot is fixed and with p when that jot occurs with a probability p, and examine the errors extensively as in Fig. XVI-4. Here we see that the frequency of the erroneous formulas is $p(1-p)$, and the actual error is a deficit of a jot at the left in each faulty state of the net, i.e., in one case each. Hence we may write for the best of stable nets $P_2 = 2^{-2} p(1-p)$. The factors p and $(1-p)$ are to be expected in the errors of any net which is logically stable, for the errors are zero when $p = 0$ or $p = 1$. No stable net is more reliable.

No designer of a computing machine would specify a neuron that was wrong more than half of the time; for he would regard it as some other neuron wrong less than half of the time; but in these most useful of logically stable circuits, it makes no difference which way he regards it, for they are symmetrical in p and $1 - p$. At $p = 1/2$, the frequency of errors is maximal and is $P_2 = 2^{-2}\, 1/2(1 - 1/2) = 1/16$, which is twice as reliable as its component neurons for which $P_1 = 2^{-2}\, 1/2 = 1/8$.

Among logically unstable circuits the most reliable can be constructed to secure fewer errors than the stable whenever $p < 1/2$. The best are like that of Fig. XVI-5. The errors here are concentrated in the two least frequent states and in only one of the four cases. Hence $P_2 = 2^{-2} p^2$.

When $\delta = 3$, the redundancy, $R = \left(2^{2^{\delta}}\right)^{\delta}$, provides so many more best stable and best unstable nets that the numbers become unwieldy. There are $\left(2^{2^{\delta}}\right)^{\delta+1}$ nets for functions of the second rank each made of 4 neurons to be selected from 8! diagrams with 9 thresholds apiece. Yet it is clear that the best stable and unstable for $\delta > 2$ are better

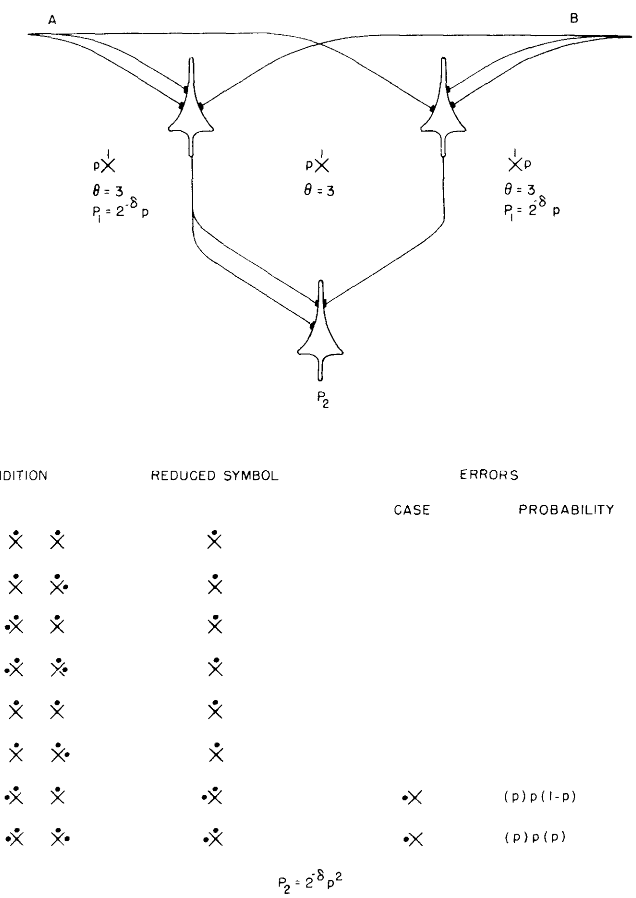

Fig. XVI-5. A best unstable net.

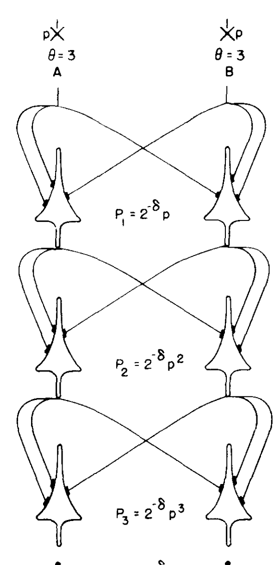

Fig. XVI-6. Unstable improvement net.

than those for $\delta = 2$ only in the factor $2^{-\delta}$ for error in a single case.

Further improvement requires the construction of nets to repeat the improvement of the first net and, for economy, the number of neurons should be a minimum. For functions of δ arguments each neuron has inputs from δ neurons. Hence the width of any rank is δ, except the last, or output, neuron. If n be the number of ranks, then the number of neurons, N, is $\delta(n-1)+1$.

Figure XVI-6 shows how to construct one of the best possible nets for the unstable ways of securing improvement with two output neurons as inputs for the next rank. The formulas are selected to exclude common errors in the output neurons on any occasion. In these, the best of unstable nets, the errors of the output neurons are $P_n = 2^{-\delta}p^n$.

[Whether we are interested in shifts of threshold or in noisy signals it is proper to ask what improvement is to be expected when two or three extra jots appear in the symbols. With our neurons a second extra jot appears only if the first has appeared, and a third only if the second. If the probability p of an extra jot is kept constant, the probability of two extra jots is p^2 and of three is p^3. Examination of the net in Fig. XVI-6 shows that $P_2 < P_1$, if $p + p^2 + p^3 < 0.15$ or $p < 0.13$. To match Gaussian noise the log of successive p's should decrease as $(\Delta\theta)^2$, or 1, p, p^4, p^9, giving $P_2 < P_1$ for $p < 0.25$. The remaining errors are always so scattered as to preclude further improvement in any subsequent rank.]

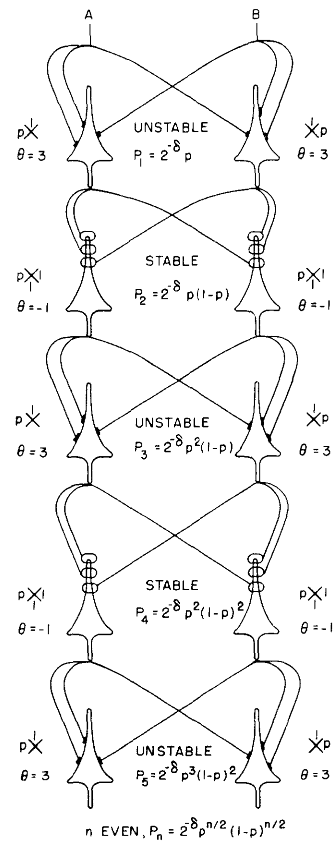

Fig. XVI-7. Alternating improvement net.

When common shifts of θ are to be expected, or all we know is $0 < p < 1$, a greater improvement is obtained by alternating stable and unstable nets as in Fig. XVI-7, selected to exclude common errors in its output neurons. The improvement for n even is

$$P_n = 2^{-\delta}\, p^{n/2}\,(1-p)^{n/2}$$

and the errors to be expected are

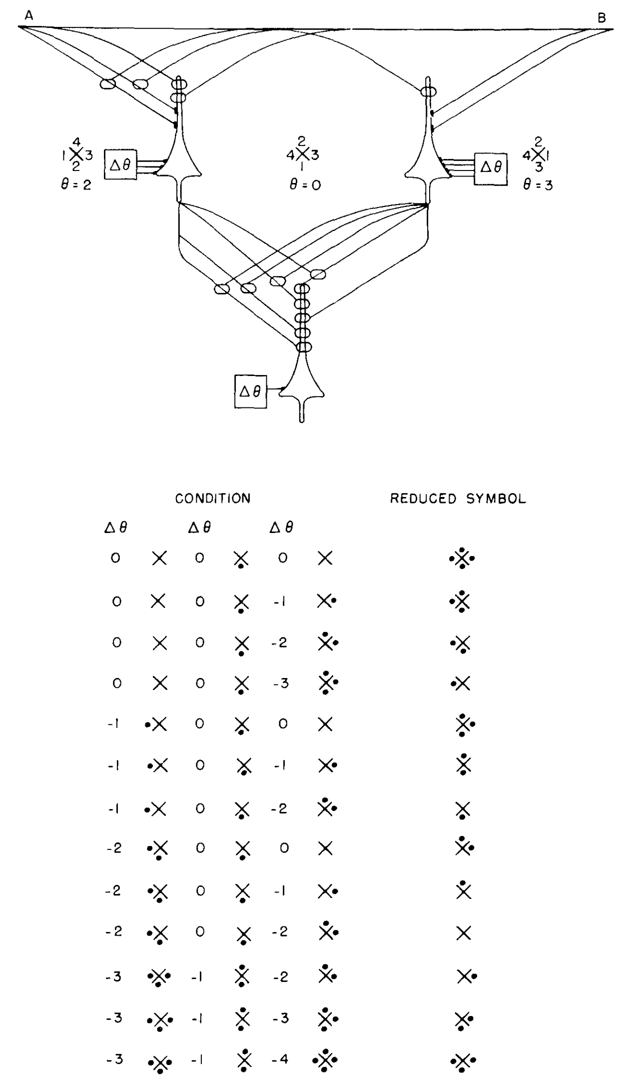

Fig. XVI-8. Flexible unstable net.

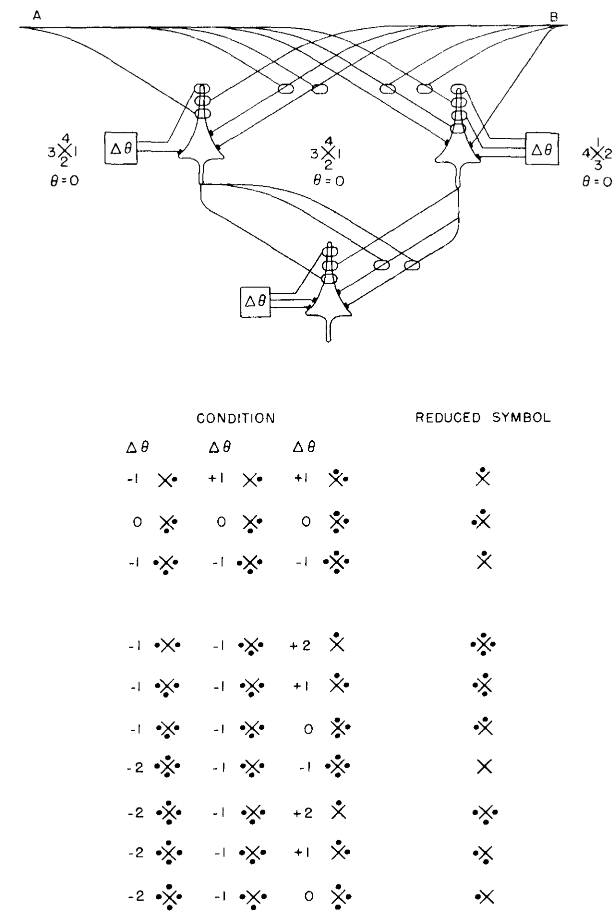

Fig. XVI-9. Flexible stable net.

$$2^{-\delta} \int_0^1 p^{n/2} (1-p)^{n/2}\, dp = 2^{-\delta} \frac{\left(\frac{n}{2}!\right)^2}{(n+1)!}$$

which is less than

$$2^{-\delta} \frac{\left(\frac{n}{2}!\right)^2}{(n+1)!} \cdot \frac{n+1}{n}$$

and hence a minimum for any proportion of stable and unstable nets.

None of these nets increases reliability in successive ranks under von Neumann's condition that neurons fire or fail with probability p regardless of input; but they are more realistic and more interesting neurons.

Despite the increase in reliability of a net for a function of the second rank these nets can be compelled to compute as many as 13 of the 16 reduced functions by altering the thresholds of the neurons by signals from other parts of the net, as in Fig. XVI-8, and even logically stable nets for triples of functions can be made to compute 8 of the 16, as in Fig. XVI-9.

Thus these appropriate neurons have served our purpose in answering the questions of logical stability under common shift of threshold and of securing reliable performance from unreliable components without losing the flexibility of functions computed by nets composed of them.

W. S. McCulloch

References

1. J. von Neumann, Fixed and random logical patterns and the problem of reliability, American Psychiatric Association, Atlantic City, May 12, 1955.
2. W. Ross Ashby, Design For a Brain (John Wiley and Sons, Inc., New York, 1st Ed., 1952).
3. J. von Neumann and O. Morgenstern, Theory of Games and Economic Behavior (Princeton University Press, Princeton, N.J., 1st Ed., 1944).
4. J. von Neumann, Probabilistic Logics and the Synthesis of Reliable Organisms From Unreliable Components (Automata Studies, Edited by C. E. Shannon and J. McCarthy, Princeton University Press, Princeton, N.J., 1956).
5. W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. and Biophys. 5, 115 (1943).
6. W. Pitts and W. S. McCulloch, How we know universals. The perception of auditory and visual forms, Bull. Math. and Biophys. 9, 127 (1947).
7. K. Menger, Analysis without variables, J. Symbol. Logic 11, No. 1, 30-31 (1946).
8. K. Menger, Algebra of Analysis (Notre Dame Mathematical Lectures, No. 3, University of Notre Dame, Notre Dame, Ind., 1944).
9. K. Menger, General algebra of analysis, Reports of a Mathematical Colloquium, Series 2, issue 7, p. 46 (1946).
10. W. S. McCulloch, Machines That Know and Want (Brain and Behavior, A Symposium, edited by W. C. Halstead; Comparative Psychology Monographs 20: No. 1: p. 39, University of California Press, Berkeley and Los Angeles, 1950).