

## XV. NEUROPHYSIOLOGY\*

W. S. McCulloch  
E. M. Duchane

R. C. Gesteland  
J. Y. Lettvin

W. H. Pitts  
P. D. Wall

### A. STABLE, RELIABLE, AND FLEXIBLE NETS OF UNRELIABLE FORMAL NEURONS (A revised and expanded<sup>†</sup> version of "Three of von Neumann's Biological Questions," Quarterly Progress Report, Oct. 15, 1957).

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 introduction of nonlinear components, (b) the negative feedback of reflexive mechanisms, (c) 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 sensitivity of thresholds or incoming signals to components that are more appropriate to components 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 a 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 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.

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, ' $\sim$ ' for 'not'; ' $\cdot$ ' for 'both'; ' $\vee$ ' for 'or'; ' $\supset$ ' for 'implies'; and ' $\equiv$ ' for 'if and only if', shorten the text but require conventions and rearrangements in order to avoid ambiguities. Economy requires one symbol for each of the sixteen logical functions of two

---

\* 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.

<sup>†</sup> For the correction of our previous publication, Quarterly Progress Report, Oct. 15, 1957, pp. 129-138, you are indebted principally to Dr. John Abbott, of Massachusetts General Hospital, Boston, Massachusetts. For any new errors, hold me guilty. W. S. McC.

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 '×' it is convenient 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. Thus we write  $A \times B$  for contradiction;  $A \cdot \times B$  for  $A \cdot \sim B$ ;  $A \dot{\times} B$  for  $A \cdot B$ ;  $A \times \cdot B$  for  $B \cdot \sim A$ ;  $A \times \cdot B$  for  $\sim A \cdot \sim B$ ;  $A \cdot \dot{\times} B$  for  $A \cdot (B \vee \sim B)$ ;  $A \dot{\times} \cdot B$  for  $(A \vee \sim A) \cdot B$ ;  $A \times \cdot B$  for  $\sim A \cdot (B \vee \sim B)$ ;  $A \cdot \times \cdot B$  for  $(A \vee \sim A) \cdot \sim B$ ;  $A \cdot \times \cdot B$  for  $(A \cdot \sim B) \vee (\sim A \cdot B)$ ;  $A \dot{\times} B$  for  $A \equiv B$ ;  $A \times \cdot B$  for  $B \supset A$ ;  $A \cdot \times \cdot B$  for  $A \vee B$ ;  $A \dot{\times} \cdot B$  for  $A \supset B$ ;  $A \times \cdot B$  for  $\sim(A \cdot B)$ ; and  $A \dot{\times} \cdot B$  for tautology. The  $\times$  may be regarded as part of the familiar Venn symbols for two classes of events of which  $A$  and  $B$  are true, respectively. Similar symbols can therefore be made, from Venn symbols, for functions of more than two arguments. Each additional line must divide every pre-existing area into two parts. Hence, for the number of arguments  $\delta$  there are  $2^\delta$  spaces for jots and  $2^{2^\delta}$  symbols for functions. (See Fig. XV-1.)

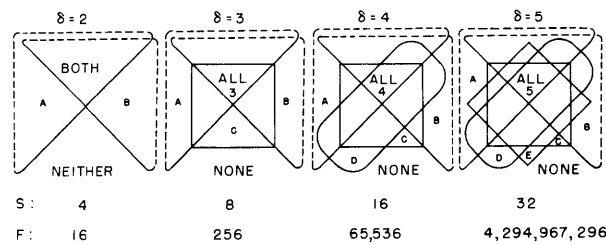


Fig. XV-1. Venn figures with spaces for all intersections of  $\delta$  classes. S is the number of spaces; F, the number of functions.

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

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

(XV. NEUROPHYSIOLOGY)

A formula of the second rank is made by inserting into the spaces for the arguments of its  $\times$  two formulas of the first rank; for example,  $(A \times B) \dot{\times} (A \times B)$ . When the two propositions of the first rank are composed of the same pair of propositions in the same order, the resulting formula of the second rank can always be equated to a formula of the first rank by putting jots into the  $\times$  for the corresponding formula of the first rank according to the following rules of reduction:

Write the equation in the form  $(\dots x_1 \dots) x_2 (\dots x_3 \dots) = (\dots x_4 \dots)$ ; wherein the  $x_j$  are chiastan symbols:

- 1) If  $x_2$  has a jot on its left, put a jot into  $x_4$  in every space where there is a jot in  $x_1$  and no corresponding jot in  $x_3$ . Thus,  $(A \dot{\times} B) \times (A \times B) = (A \times B)$
- 2) If  $x_2$  has a jot on its right, put a jot into  $x_4$  in every space where there is a jot in  $x_3$  and no corresponding jot in  $x_1$ . Thus,  $(A \times B) \times (A \dot{\times} B) = (A \times B)$
- 3) If  $x_2$  has a jot above, put a jot into  $x_4$  in every space where there is a jot in both  $x_1$  and  $x_3$ . Thus,  $(A \dot{\times} B) \dot{\times} (A \times B) = (A \times B)$
- 4) If  $x_2$  has a jot below, put a jot into  $x_4$  in every space that is empty in both  $x_1$  and  $x_3$ . Thus,  $(A \times B) \times (A \times B) = (A \dot{\times} B)$

If there is more than one jot in  $x_2$  apply the foregoing rules seriatim until all jots on  $x_2$  have been used. Put no other jots into  $x_4$ .

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 no other formulas are used in this article, the letters A and B are omitted, and positions, left and right, replace parentheses.

In formulas 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 the symbols proposed for functions of three arguments, the error will occur in only one of the eight cases, and, in general, for functions of  $\delta$  arguments, in one of  $2^\delta$  cases. If  $p$  is the probability of the erroneous jot and  $P$  the probability of error produced,  $P = 2^{-\delta} p$ . In empirical examples the relative frequency of the case in question as a matter of fact replaces the logical probability.

In formulas for the second rank there are three  $\times$ 's. 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  $16^3/16 = 16^2$ . For functions of  $\delta$  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 formal 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, as we have shown (11, 12), 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. XV-2 are merely suggested by the anatomy of the nervous system. 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 shows that it 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,  $\theta$ , measured in steps of excitation, and it emits a signal when the excitation it receives is equal to or greater than  $\theta$ . 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  $\times$ 's and jots beginning with none and adding one at a time as  $\theta$  decreases until all four have appeared in the sequence noted in the legend for its diagram in Fig. XV-2. These are the simplest diagrams fulfilling the requirement. All simpler diagrams are degenerate, since they either fail to add one jot or else add more than one jot for some step in  $\theta$ . 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  $\times$  as a change on the threshold of an appropriate neuron.

Any formula of the second rank is realized by a net of three neurons each of whose thresholds is specified; for example, see Fig. XV-3. The formula can be reduced to one of the first rank whose  $\times$  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  $\times$  but the reduced formula is unaltered, the net is called "logically stable."

The redundancy of formulas of the second rank provides us with many examples of pairs of formulas and even triples of formulas that reduce to the same formula of the first rank and that can be made from one another by common addition or omission of one jot in each  $\times$ , and the diagrams of Fig. XV-2 enable us to realize them all in several ways: For example, there are 32 triples of formulas and 64 logically stable nets for every reduced formula with a single jot. Even nets of degenerate diagrams enjoy some logical stability; for example  $\times \times \cdot \times = \dot{\times}$  goes to  $\cdot \times \cdot \times \cdot \times = \dot{\times}$ .

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

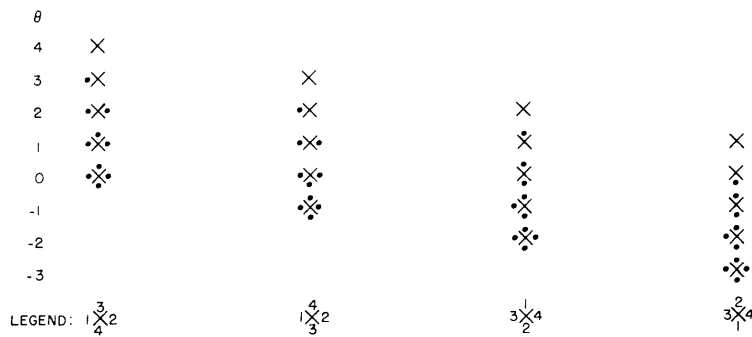
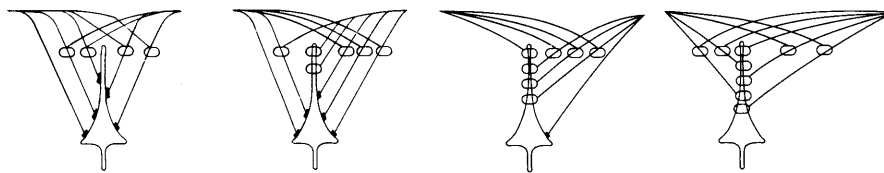
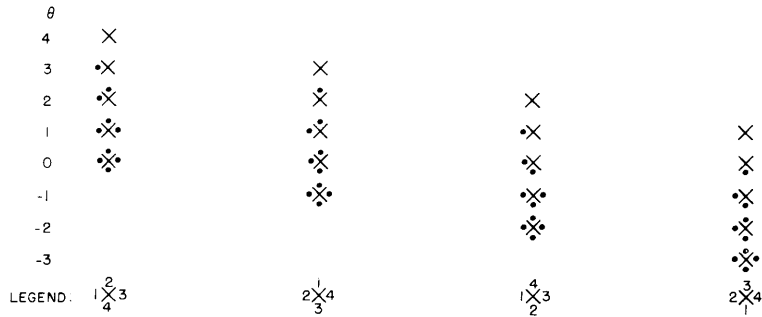
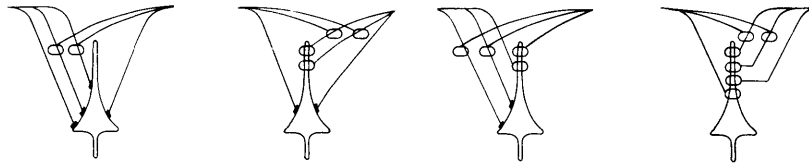
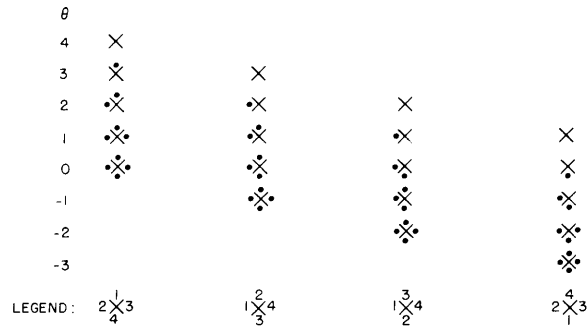
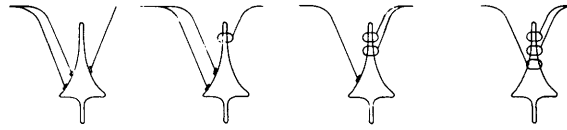


Fig. XV-2. Synaptic diagrams for  $\times$  appearing before  $\times$ .

coma, for no formula is significant or its net stable with a shift of  $\theta$  that compels the output neuron to compute tautology or contradiction. The net of Fig. XV-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 independently in their neurons and the most reliable make some errors; for example, the net of Fig. XV-4.

To include independent shifts, let our chiastan symbols be modified by replacing a jot with 1 when the jot is never omitted and with  $p$  when that jot occurs with a probability  $p$ , and examine the errors extensively as in Fig. XV-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 the reduced formula 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 produced by 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 error 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. XV-5. The errors here are concentrated in the two least frequent states and in only one of the four cases. Hence  $P_2 = 2^{-2} p^2$ .

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  $\delta$  arguments each neuron has inputs from  $\delta$  neurons. Hence the width of any rank is  $\delta$ , 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 XV-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 nondegenerate diagrams for 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

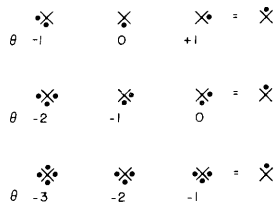
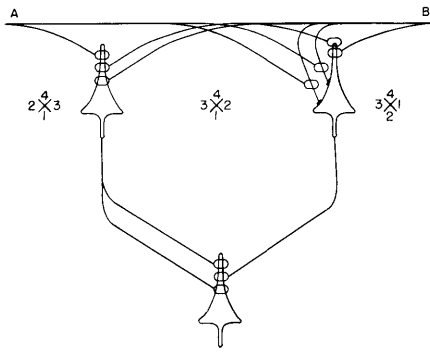
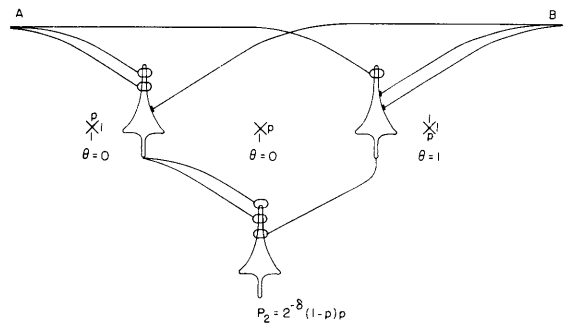
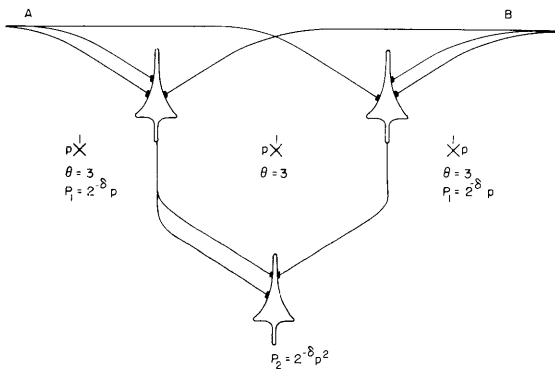


Fig. XV-3. A logically stable net for X.



| CONDITION | REDUCED SYMBOL | ERRORS |             |
|-----------|----------------|--------|-------------|
|           |                | CASE   | PROBABILITY |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              | X      | (1-p)p(1-p) |
| X X X     | X              | X      | (1-p)p(p)   |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              |        |             |

Fig. XV-4. A best stable net for X.



| CONDITION | REDUCED SYMBOL | ERRORS |             |
|-----------|----------------|--------|-------------|
|           |                | CASE   | PROBABILITY |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              |        |             |
| X X X     | X              | X      | (p)p(1-p)   |
| X X X     | X              | X      | (p)p(p)     |

Fig. XV-5. A best unstable net for X.

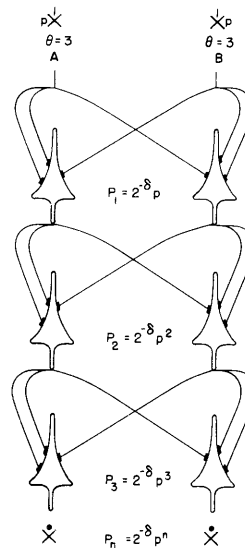


Fig. XV-6. Unstable improvement net for X.

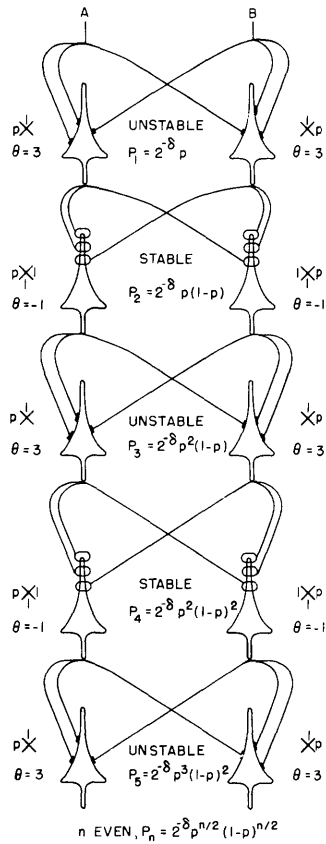


Fig. XV-7. Alternating improvement net for  $\bar{X}$ .

When  $\delta = 3$ , the redundancy,  $R = (2^{2^\delta})^\delta$ , provides so many more best stable and best unstable nets that the numbers become unwieldy. There are  $(2^{2^\delta})^{\delta+1}$  nets for functions of the second rank each made of 4 neurons to be selected from  $8!$  diagrams with 9 thresholds apiece. Formerly I said it was clear that the best stable and unstable nets for  $\delta > 2$  are better than those for  $\delta = 2$  only in the factor  $2^{-\delta}$  for error in a single case. That is only true if the nets are composed of nondegenerate diagrams alone. With neurons  $\delta = 3$ , a single degenerate diagram for the output neuron permits the construction of a logically stable net with  $P_2 = 0$ , even with independent shifts of  $\theta$  sufficient to alter the logical function computed by every neuron, as seen in Fig. XV-8. The same degenerate diagram for the  $\delta = 3$  output neuron receiving inputs from three nondegenerate  $\delta = 2$  neurons, selected to make but one error in each case, is likewise stable and has an error-free output despite independent shifts of  $\theta$ , as is seen in Fig. XV-9.

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 interesting neurons. They are also more realistic. In the best controlled

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. XV-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.]

When common shifts of  $\theta$  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. XV-7, selected to exclude common errors in its output neurons. For  $n$  even

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

and the expected error is

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

which is less than with any other compositions of  $\delta = 2$  nondegenerate diagrams.



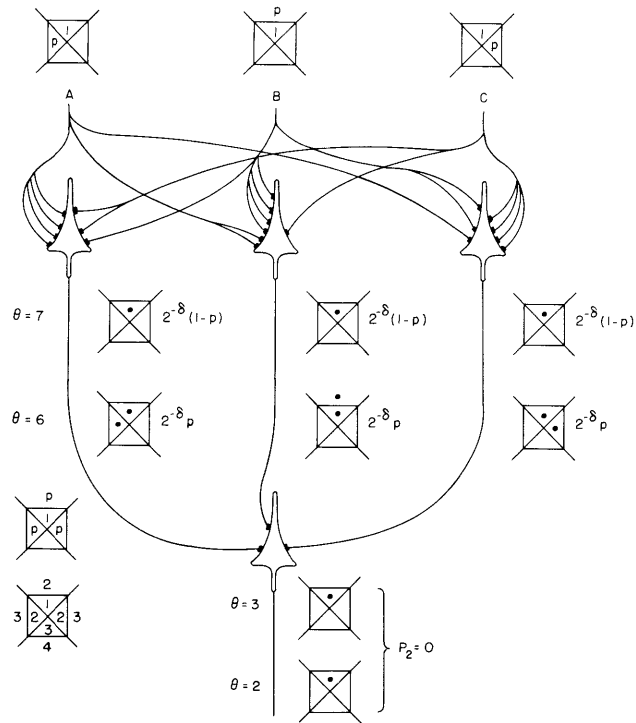


Fig. XV-8. Input  $\delta = 3$ , output degenerate  $\delta = 3$  neuron for  $\dot{\times}$ .

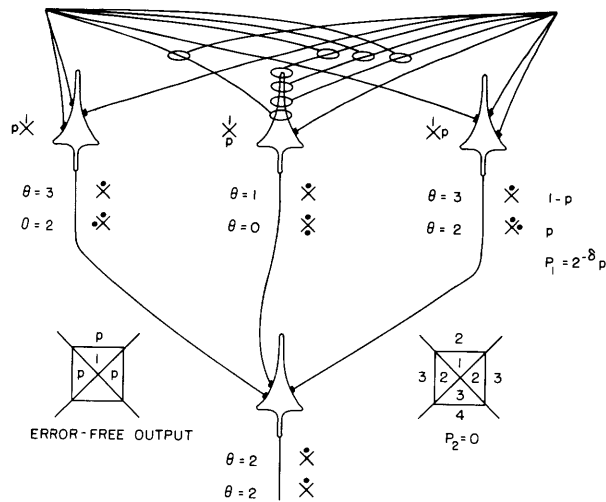
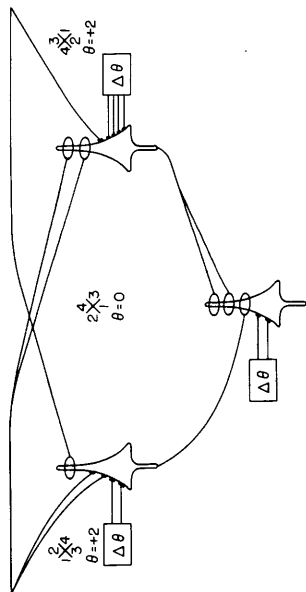


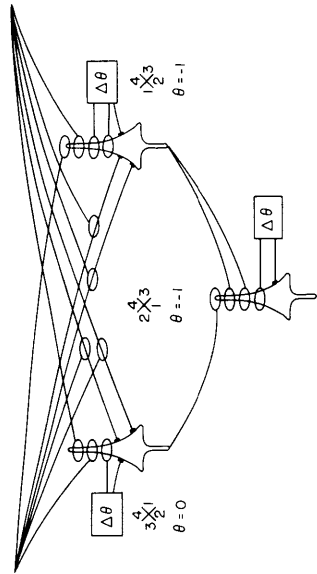
Fig. XV-9. Input  $\delta = 2$ , output degenerate  $\delta = 3$  neuron for  $\dot{\times}$ .



14 REDUCED SYMBOLS

| CONDITION | $\Delta\theta$ | $\Delta\theta$ | $\Delta\theta$ |
|-----------|----------------|----------------|----------------|
| 0         | +              | +              | +              |
| 0         | +              | +              | +              |
| -1        | +              | +              | +              |
| -1        | +              | +              | +              |
| -2        | +              | +              | +              |
| 0         | +              | +              | +              |
| 0         | +              | +              | +              |
| -1        | +              | +              | +              |
| -1        | +              | +              | +              |
| 0         | +              | +              | +              |
| 0         | +              | +              | +              |
| -1        | +              | +              | +              |
| -1        | +              | +              | +              |
| -2        | +              | +              | +              |
| -2        | +              | +              | +              |

Fig. XV-10. Flexible net, always unstable.



9 REDUCED SYMBOLS

| CONDITION | $\Delta\theta$ | $\Delta\theta$ | $\Delta\theta$ |
|-----------|----------------|----------------|----------------|
| +         | +              | +              | +              |
| 0         | +              | +              | +              |
| -1        | +              | +              | +              |
| +         | +              | +              | +              |
| 0         | +              | +              | +              |
| +         | +              | +              | +              |
| +         | +              | +              | +              |
| 0         | +              | +              | +              |
| 0         | +              | +              | +              |
| 0         | +              | +              | +              |
| 0         | +              | +              | +              |

Fig. XV-11. Flexible net, logically stable for  $\dot{X}$ .

## (XV. NEUROPHYSIOLOGY)

experiments the record of a single unit, be it cell body or axon, always indicates firing and failing to fire at near threshold values of constant excitation, whether it is excited transsynaptically or by current locally applied. At present, we do not know how much of the observed flutter of threshold is due to activity of other afferent impulses and how much is intrinsic fluctuation at the trigger point. We have not yet a reasonable guess as to its magnitude in neurons in situ, and for excised nerve we have only two estimates of the range: one, about 2 per cent; and the other, 7 per cent (13, 14). Our own measurements on a single node of Ranvier would indicate a range of 10 per cent. To discover regularities of nervous activity we have naturally avoided stimulation near threshold. Now we must measure the intrinsic jitter, for this determines both the necessity of improving circuits and the maximum number of afferent terminals that can be discriminated. Eventually we will have to take into account the temporal course of excitation and previous response, for every impulse leaves a wake of changing threshold along its course.

Despite the increase in reliability of a net for a function of the second rank some of these nets can be compelled to compute as many as 14 of the 16 reduced formulas by altering the thresholds of the neurons by means of signals from other parts of the net, as in Fig. XV-10, and even some logically stable nets for triples of formulas can be made to compute 9 of the 16, as in Fig. XV-11. Even this does not exhaust the redundancy of these nets, for both of them can be compelled to compute many of these functions in several ways.

The diagrams of Fig. XV-2 were drawn to insure a change in function for every step in  $\theta$ . Actual neurons have more redundant connections. We are examining how to use this redundancy to design reliable nets the details of whose connections are subject to statistical constraints alone. This is important because our genes cannot carry enough information to specify synapsis precisely.

For the moment, it is enough that these appropriate, formal neurons have demonstrated logical stability under common shift of threshold and have secured reliable performance by nets of unreliable components without loss of flexibility.

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).
11. W. S. McCulloch, B. Howland, J. Y. Lettvin, W. Pitts, and P. D. Wall, Reflex inhibition by dorsal root interaction, *J. Neurophysiol.* 18, 1-17 (1955).
12. W. S. McCulloch, P. D. Wall, J. Y. Lettvin, W. H. Pitts, Factors limiting the maximum impulse transmitting ability of an afferent system of nerve fibers, *Information Theory*, Third London Symposium, edited by C. Cherry (Academic Press, Inc., New York, 1956), p. 329.
13. E. A. Blair and J. A. Erlanger, A comparison of the characteristics of axons through their individual electrical responses, *Am. J. Phys.* 106, 524-564 (1933).
14. C. Pecher, La fluctuation d'excitabilité de la fibre nerveuse, *Arch. Internat. de Physiol.* 49, 129-152 (1939).