JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 21, 643–694 (1968)

A Prediction Theory for Some Nonlinear Functional-Differential Equations I. Learning of Lists

STEPHEN GROSSBERG

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Submitted by Gian-Carlo Rota

PART I

1. INTRODUCTION

In this paper, we study some systems of nonlinear functional-differential equations of the form

$$\dot{X}(t) = AX(t) + B(X_t)\, X(t-\tau) + C(t), \qquad t \geqslant 0, \tag{1}$$

which were introduced in Grossberg ([1], [2], [3]). We will choose (1) so that $X = (x_1, x_2, ..., x_n)$ is nonnegative, $B(X_t) = \| B_{ij}(t) \|$ is a matrix of nonnegative and nonlinear functionals of $X(w)$ evaluated at all past times $w \in [-\tau, t]$, and $C = (I_1, I_2, ..., I_n)$ is a known nonnegative and continuous input function. We will show that for appropriate choices of A, B, and C, ratios such as

$$X_k(t) = x_k(t) \left(\sum_{m=1}^{n} x_m(t) \right)^{-1}$$

have limits as $t \to \infty$, for all $j, k = 1, 2, ..., n$.

For these choices of A, B, and C, we will be able to interpret (1) as a prediction theory. The goal of this theory is to discuss the prediction of individual events, in a fixed order, and at prescribed times. The theory is not homogeneous in time. A system which produces random predictions at $t = 0$ can be gradually transformed into a system whose predictions become deterministic as $t \to \infty$. Similarly, a system which produces deterministic predictions at $t = 0$ can be gradually transformed into a system whose predictions become random as $t \to \infty$. The factor which primarily determines if a system becomes random or deterministic in its predictions as $t \to \infty$ is the system's input function $C(t)$. $C(t)$ is the "environment" or "experience" of the system, and we will make precise the statement that these systems "adapt to their environment" or "learn from experience."

Our systems can also be interpreted as cross-correlated flows on networks, or as deformations of probabilistic graphs. They often have the property that the average input

$$I(t) = \frac{1}{n} \sum_{k=1}^{n} I_k(t)$$

is related to the average output

$$x(t) = \frac{1}{n} \sum_{k=1}^{n} x_k(t)$$

through a system of *linear* difference-differential equations. This property is crucial to our proofs. Another important property is the nonnegativity of initial data. When mixtures of positive and negative initial data are chosen, the results are not true in general.

2. The Systems and Their Basic Properties

A and B are chosen in the following way. Let us be given any positive integer n; any real numbers, α, u, $\beta > 0$, and $\tau \geqslant 0$; and any $n \times n$ semistochastic matrix $P = \| p_{ij} \|$ (i.e., $p_{ij} \geqslant 0$ and $\sum_{k=1}^{n} p_{ik} = 0$ or 1). Then we let

$$\left.\begin{aligned} \dot{x}_i(t) &= -\alpha x_i(t) + \beta \sum_{k=1}^{n} x_k(t-\tau)\, y_{ki}(t) + I_i(t), && (2) \\ y_{jk}(t) &= p_{jk} z_{jk}(t) \left(\sum_{m=1}^{n} p_{jm} z_{jm}(t) \right)^{-1}, && (3) \\ \text{and} \\ \dot{z}_{jk}(t) &= [-u z_{jk}(t) + \beta x_j(t-\tau)\, x_k(t)]\, \theta(p_{jk}), && (4) \end{aligned}\right\} *$$

where

$$\theta(p) = \begin{cases} 1 & \text{if} \quad p > 0 \\ 0 & \text{if} \quad p \leqslant 0 \end{cases}$$

and $i, j, k = 1, 2, \ldots, n$. If for example all p_{jk} are positive, then in (1),

$$A_{ij}(t) = -\alpha \delta_{ij}$$

and

$$B_{ij}(t) = \frac{p_{ji}\left[z_{ji}(0) + \int_0^t e^{uv} x_j(v - \tau)\, x_i(v)\, dv\right]}{\sum_{m=1}^{n} p_{jm}\left[z_{jm}(0) + \int_0^t e^{uv} x_j(v - \tau)\, x_m(v)\, dv\right]}.$$

All of our results require that the initial data of (*) be *nonnegative*. We also require the initial data to be continuous and for convenience let $z_{jk}(0) > 0$ iff $p_{jk} > 0$. When we say the initial data is chosen "arbitrarily," we will always mean "arbitrarily subject to these constraints."

The following theorem guarantees that (*) makes sense when its initial data are chosen in this way.

THEOREM 1. *Let* (*) *be given with arbitrary initial data. Then the solution of* (*) *exists and is unique, continuously differentiable, and nonnegative in* $(0, \infty)$. *If the initial data of a given variable* x_i *or* z_{jk} *is positive, then this variable is positive in* $(0, \infty)$.

PROOF. (*) can be written in vector form as

$$\dot{U}(t) = f(t, U(t), U(t - \tau)) \tag{*}$$

with

$$U = (x_1, x_2, \ldots, x_n, z_{11}, z_{12}, \ldots, z_{n,n-1}, z_{nn}),$$

$$f = (f_1, f_2, \ldots, f_n, f_{11}, f_{12}, \ldots, f_{n,n-1}, f_{nn}),$$

$$f_i = -\alpha x_i + \beta \sum_{k=1}^{n} x_k(t - \tau)\, p_{ki} z_{ki} \left(\sum_{m=1}^{n} p_{km} z_{km}\right)^{-1} + I_i,$$

and

$$f_{jk} = [-u z_{jk} + \beta x_j(t - \tau)\, x_k]\, \theta(p_{jk}).$$

Let $\tau = 0$. Then $\dot{U}(t) = g(t, U(t))$, where $g(t, w) = f(t, w, w)$. By the continuity of g, a solution $U(t)$ exists in an interval with 0 as its left-hand endpoint. If moreover,

$$| g(t, U^{(1)}) - g(t, U^{(2)} | \leqslant k(t) \mid U^{(1)} - U^{(2)} |$$

for some continuous function $k(t)$ and any two solutions $U^{(1)}$ and $U^{(2)}$, then this interval is $(0, \infty)$ and the solution is unique and continuously differentiable [5]. First we show that such a $k(t)$ exists if all x_i and z_{jk} are nonnegative. The only terms for which this is not obvious are the terms

$$x_j p_{jk} z_{jk} \left(\sum_{m=1}^{n} p_{jm} z_{jm}\right)^{-1}$$

We use nonnegativity to estimate x_j above by a continuous function $m(t)$. By nonnegativity, $\dot{x}_i \geqslant -\alpha x_i$ and $\dot{z}_{jk} \geqslant -u z_{jk}$, or $x_i(t) \geqslant e^{-\alpha t} x_i(0)$ and $z_{jk}(t) \geqslant e^{-ut} z_{jk}(0)$. Thus nonnegativity implies positivity if the initial data is positive. In particular,

$$\sum_{m=1}^{n} p_{jm} z_{jm}(t) \geqslant e^{-ut} \sum_{m=1}^{n} p_{jm} z_{jm}(0) \begin{cases} = 0 & \text{if} \quad \sum_{m=1}^{n} p_{jm} = 0 \\ > 0 & \text{if} \quad \sum_{m=1}^{n} p_{jm} = 1. \end{cases}$$

This implies

$$\sum_{m=1}^{n} y_{jm}(t) = \sum_{m=1}^{n} p_{jm} = 0 \qquad \text{or} \qquad 1, \tag{5}$$

from which we find that

$$\dot{x} \leqslant (\beta - \alpha)\, x + I,$$

where

$$x = \frac{1}{n} \sum_{k=1}^{n} x_k \qquad \text{and} \qquad I = \frac{1}{n} \sum_{k=1}^{n} I_k\,,$$

or

$$x(t) \leqslant \frac{1}{n}\, m(t)$$

where

$$m(t) = n e^{(\beta-\alpha)t} \left[x(0) + \int_0^t e^{(\alpha-\beta)v} I(v)\, dv \right].$$

By nonnegativity,

$$x_j(t) \leqslant nx(t) \leqslant m(t).$$

We can now prove the required Lipschitz condition. Obviously

$$\left| \frac{x_j^{(1)} p_{jk} z_{jk}^{(1)}}{\sum_{m=1}^{n} p_{jm} z_{jm}^{(1)}} - \frac{x_j^{(2)} p_{jk} z_{jk}^{(2)}}{\sum_{m=1}^{n} p_{jm} z_{jm}^{(2)}} \right|$$

$$\leqslant | x_j^{(1)} - x_j^{(2)} | + m(t) \left| \frac{p_{jk} z_{jk}^{(1)}}{\sum_{m=1}^{n} p_{jm} z_{jm}^{(1)}} - \frac{p_{jk} z_{jk}^{(2)}}{\sum_{m=1}^{n} p_{jm} z_{jm}^{(2)}} \right|.$$

It therefore suffices to show that

$$p_{jk}\left|\frac{z_{jk}^{(1)}}{\sum_{m=1}^{n} p_{jm}z_{jm}^{(1)}}-\frac{z_{jk}^{(2)}}{\sum_{m=1}^{n} p_{jm}z_{jm}^{(2)}}\right|\leqslant h(t)\sum_{m=1}^{n}|z_{jm}^{(1)}-z_{jm}^{(2)}|$$

for some continuous $h(t)$. When $\sum_{m=1}^{n} p_{jm}=0$, the choice $h(t)=0$ suffices. Suppose $\sum_{m=1}^{n} p_{jm}=1$. Then

$$p_{jk}\left|\frac{z_{jk}^{(1)}}{\sum_{m=1}^{n} p_{jm}z_{jm}^{(1)}}-\frac{z_{jk}^{(2)}}{\sum_{m=1}^{n} p_{jm}z_{jm}^{(2)}}\right|$$

$$=p_{jk}\left|\frac{(z_{jk}^{(1)}-z_{jk}^{(2)})\sum_{m=1}^{n} p_{jm}z_{jm}^{(2)}+z_{jk}^{(2)}\sum_{m=1}^{n} p_{jm}(z_{jm}^{(2)}-z_{jm}^{(1)})}{\left(\sum_{m=1}^{n} p_{jm}z_{jm}^{(1)}\right)\left(\sum_{m=1}^{n} p_{jm}z_{jm}^{(2)}\right)}\right|$$

$$\leqslant\frac{|z_{jk}^{(1)}-z_{jk}^{(2)}|+\sum_{m=1}^{n} p_{jm}|z_{jm}^{(2)}-z_{jm}^{(1)}|}{\sum_{m=1}^{n} p_{jm}z_{jm}^{(1)}}$$

$$\leqslant\frac{e^{ut}}{\sum_{m=1}^{n} p_{jm}z_{jm}^{(1)}(0)}\left(|z_{jk}^{(1)}-z_{jk}^{(2)}|+\sum_{m=1}^{n}|z_{jm}^{(1)}-z_{jm}^{(2)}|\right)$$

$$\leqslant\frac{2e^{ut}}{\sum_{m=1}^{n} p_{jm}z_{jm}^{(1)}(0)}\sum_{m=1}^{n}|z_{jm}^{(1)}-z_{jm}^{(2)}|.$$

Letting

$$h(t)=2e^{ut}\left(\sum_{m=1}^{n} p_{jm}z_{jm}^{(1)}(0)\right)^{-1}$$

completes the proof when $\tau=0$ except for the demonstration that x_i and z_{jk} are nonnegative.

By (4) and the nonnegativity of initial data, $z_{jk}(t)$ cannot become negative until either $x_j(t)$ or $x_k(t)$ becomes negative. Otherwise if $z_{jk}(t)$ is zero at $t = T_0$, then

$$\dot{z}_{jk}(T_0) \geqslant x_j(T_0)\, x_k(T_0) \geqslant 0.$$

Let $t = T_1$ be the first zero of any function $x_i(t)$. Suppose in particular that $x_1(T_1) = 0$. Then by (2) and the nonnegativity of $C(t)$,

$$\dot{x}_1(T_1) = \beta \sum_{k=1}^{n} x_k(T_1)\, y_{k1}(T_1) + I_1(T_1) \geqslant 0.$$

x_1 can therefore never become negative, and all solutions are nonnegative. This completes the proof when $\tau = 0$.

Suppose $\tau > 0$. The existence of a solution of (*) follows by a standard "step-by-step" construction in each interval of the form $(n\tau, (n+1)\tau]$, $n = 0, 1,...$ [7]. To prove the remaining assertions, it suffices to show that $|f(\xi_1, \eta) - f(\xi_2, \eta)| \leqslant k(t)\, |\xi_1 - \xi_2|$ for every η, and this can be done just as in the case $\tau = 0$. The proof is therefore complete.

Theorem 1 implies a property of averages of the inputs I_i and the outputs x_i that is used repeatedly in proving our results. To state this property, we inductively define a sequence of subsets $S(r)$ and $T(r)$ of $\{1, 2,..., n\}$ by

$$S(r) = \Big\{k : \sum_{i \in S(r-1)} p_{ki} = 1\Big\}$$

and

$$T(r) = \Big\{k : \sum_{i \in S(r-1)} p_{ki} = 0\Big\}, \qquad r = 1, 2,..., k,$$

where $S(0) = \{1, 2,..., n\}$ and k is the least positive integer such that each $S(r)$ and $S(r)\backslash S(r-1)$ is nonempty, $r = 1, 2,..., k-1$. We also let

$$x_{(r)} = \sum_{i \in S(r)} x_i \qquad \text{and} \qquad I_{(r)} = \sum_{i \in S(r)} I_i .$$

COROLLARY 1. *The vectors* $V = (x_{(0)}, ..., x_{(k-1)})$ *and* $W = (I_{(0)}, ..., I_{(k-1)})$ *obey a linear equation*

$$\dot{V}(t) = -\alpha V(t) + \beta D V(t - \tau) + W(t) \tag{6}$$

iff

$$S(r) \cup T(r) = S(0), \qquad r = 1, 2,..., k.$$

When $S(k) = S(k-1)$,

$$D = \begin{pmatrix} 0 & 1 & 0 & . & . & . & . \\ 0 & 0 & 1 & 0 & . & . & . \\ \vdots & & & & & & \\ 0 & 0 & . & . & . & 0 & 1 \\ 0 & 0 & . & . & . & 0 & 1 \end{pmatrix}.$$

When $S(k) = \phi$,

$$D = \begin{pmatrix} 0 & 1 & 0 & . & . & . & . \\ 0 & 0 & 1 & 0 & . & . & . \\ \vdots & & & & & & \\ 0 & 0 & . & . & . & 0 & 1 \\ 0 & 0 & . & . & . & 0 & 0 \end{pmatrix}.$$

If moreover P is stochastic (i.e., $\sum_{m=1}^{n} p_{im} = 1$ *for all* i), *then* (6) *becomes*

$$\dot{x}(t) = -\alpha x(t) + \beta x(t-\tau) + I(t). \tag{7}$$

PROOF. To prove sufficiency, sum (2) over all $i \in S(r)$, for any $r = 0, 1,..., k-2$. Then

$$\dot{x}_{(r)}(t) = -\alpha x_{(r)}(t) + \beta \sum_{m=1}^{n} x_m(t-\tau) \sum_{i \in S(r)} y_{mi}(t) + I_{(r)}(t).$$

Since $S(0) = S(r) \cup T(r)$,

$$x_{(r+1)}(t-\tau) = \sum_{m=1}^{n} x_m(t-\tau) \sum_{i \in S(r)} y_{mi}(t)$$

and

$$\dot{x}_{(r)}(t) = -\alpha x_{(r)}(t) + \beta x_{(r+1)}(t-\tau) + I_{(r)}(t).$$

Let $r = k-1$. If $S(k) = S(k-1) \neq \phi$, then

$$\dot{x}_{(k-1)}(t) = -\alpha x_{(k-1)}(t) + \beta x_{(k-1)}(t-\tau) + I_{(k-1)}(t). \tag{8}$$

If $S(k) = \phi$, then

$$\dot{x}_{(k-1)}(t) = -\alpha x_{(k-1)}(t) + I_{(k-1)}(t),$$

which completes the proof of sufficiency.

Necessity follows from the observation that at least one $y_{mi}(t)$ is not summed to 0 or 1 in $\sum_{m=1}^{n} x_m(t-\tau) \sum_{i \in S(r)} y_{mi}(t)$ if $S(0) \neq S(r) \cup T(r)$, and thus the system is nonlinear, If P is stochastic, then $S(1) = S(0)$ and by (8),

$$\dot{x}_{(0)}(t) = -\alpha x_{(0)}(t) + \beta x_{(0)}(t - \tau) + I_{(0)}(t). \tag{9}$$

Dividing both sides of (9) by n gives (7).

REMARK. The vector $C = (I_1, ..., I_n)$ can be viewed as inputs fed into the machine (*) by an experimenter, and $X = (x_1, ..., x_n)$ can be viewed as the outputs produced thereby. By Corollary 1, the average input and output of the machine often obeys a simple system of *linear* equations. It is therefore natural to ask what new information the experimenter gains by studying the *nonlinear* interactions of (*) within itself. It is easily seen by nonnegativity of solutions that if $\lim_{t\to\infty} x(t) = 0$ and $u > 0$, then also

$$\lim_{t\to\infty} x_i(t) = \lim_{t\to\infty} z_{jk}(t) = 0$$

for all i, j, and k. Thus the individual variables x_i and z_{jk} need not carry any more information than the average output x as $t \to \infty$. The new information of (*) is contained, instead, in *ratios* such as

$$y_{jk}(t) = p_{jk} z_{jk}(t) \left(\sum_{m=1}^{n} p_{jm} z_{jm}(t) \right)^{-1}$$

and

$$x_{jk}(t) = p_{jk} x_k(t) \left(\sum_{m=1}^{n} p_{jm} x_m(t) \right)^{-1} \quad \text{as} \quad t \to \infty.$$

We will show, moreover, that such ratios can have a substantial effect on the actual size of the outputs $x_i(t)$, even when they have no effect on the average output $x(t)$, if the inputs $C(t)$ are properly chosen.

3. GEOMETRICAL INTERPRETATION

Before studying the ratios of (*), we give (*) a geometrical interpretation which helps to visualize and motivate our statements. This interpretation also facilitates comparison and contrast of our systems with some known biological systems to which they are similar in certain ways. This comparison will be carried out in another place.

Let G be any finite directed graph [4] with vertices $V = \{v_i : i = 1, 2, ..., n\}$ and directed edges $E = \{e_{ij} : i, j = 1, 2, ..., n\}$. To each v_i, we assign the *vertex* (or *state*) *function* $x_i(t)$, and to each e_{ij}, we assign the *edge* (or *inter-*

action) function $y_{ij}(t)$, as in Fig. 1. $x_i(t)$ is thought of as a process going on at v_i, and $y_{ij}(t)$ is though of as a process going on at the *arrowhead* of e_{ij}. With this picture in mind, (*) can be interpreted as a kind of *flow* on G [6] in the following way.

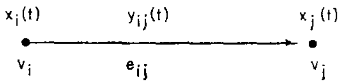

FIG. 1

(A) THE FLOW ALONG A SINGLE EDGE. At every time $t - \tau$, a quantity of size $\beta x_i(t - \tau)$ leaves vertex v_i and flows along the edge e_{ij} at a finite velocity. This quantity reaches the arrowhead of e_{ij} at time t. When $\beta x_i(t - \tau)$ reaches the arrowhead of e_{ij} at time t, it activates the process described by $y_{ij}(t)$. As a result of this activation, a total magnitude $\beta x_i(t - \tau)\, y_{ij}(t)$ is instantaneously emitted from the arrowhead and reaches vertex v_j at time t. This process is illustrated in Fig. 2.

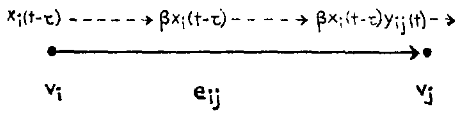

FIG. 2

(B) THE TOTAL FLOW ARRIVING AT A FIXED VERTEX. The total flow received by vertex v_j from all other vertices v_i at time t is the ***sum*** of the flows received from each vertex v_i. By (A), this flow is

$$\beta \sum_{i=1}^{n} x_i(t - \tau)\, y_{ij}(t),$$

as in Fig. 3. (2) says that the contribution of all vertices to the rate of change of the $x_j(t)$ process at v_j equals this total flow at every time t. The rate of

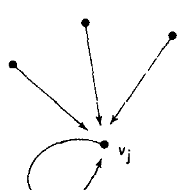

FIG. 3

change of $x_j(t)$ is also proportional to the magnitude of the input function $I_j(t)$ controlled by the experimenter, and $x_j(t)$ decays (or diverges) spontaneously at an exponential rate α.

(C) THE TOTAL FLOW LEAVING A FIXED VERTEX. By (A), the total flow received by all vertices v_i from a fixed vertex v_j at time t is

$$\beta \sum_{i=1}^{n} x_j(t-\tau)\, y_{ji}(t) = \begin{cases} 0 & \text{if} \quad \sum_{i=1}^{n} p_{ji} = 0 \\ \beta x_j(t-\tau) & \text{if} \quad \sum_{i=1}^{n} p_{ji} = 1. \end{cases}$$

Thus v_j either sends out no flow whatsoever at any time, or sends out a total flow which is proportional to its vertex function.

(D) THE FLOW IS CROSS-CORRELATED. The function $y_{ij}(t)$ which appears in the flow magnitude $\beta x_i(t-\tau)\, y_{ij}(t)$ received by v_j from v_i at time t itself depends on the vertex functions, as is obvious from (3) and (4). The term $\beta x_i(t-\tau)\, x_j(t)$ appearing in (4) has the following interpretation in terms of the flow along the edge e_{ij} . $\beta x_i(t-\tau)$ is the size of the flow received by the arrowhead of e_{ij} from v_i at time t. This arrowhead touches on v_j , whose vertex function has the value $x_j(t)$ at time t. $z_{ij}(t)$ *cross-correlates* the two quantities $\beta x_i(t-\tau)$ and $x_j(t)$ which impinge on the arrowhead at time t. That is, the rate of change of $z_{ij}(t)$ is proportional to $\beta x_i(t-\tau)\, x_j(t)$. $z_{ij}(t)$ also decays (or diverges) spontaneously at the rate u.

We form $y_{ij}(t)$ from the cross-correlating functions $z_{ik}(t)$ weighted by the coefficients p_{ik}; that is, from $p_{ik}z_{ik}(t)$. This is done by dividing $p_{ij}z_{ij}(t)$ by the sum of the functions $p_{ik}z_{ik}(t)$, $k = 1, 2,..., n$, which belong to any edge e_{ik} that faces away from v_i , as in Fig. 4. $y_{ij}(t)$ appears in the flow $\beta x_i(t-\tau)\, y_{ij}(t)$ instead of the unnormalized function $p_{ij}z_{ij}(t)$ to guarantee that the average output $x(t)$ of (*) obeys a *linear* equation, as in Corollary 1.

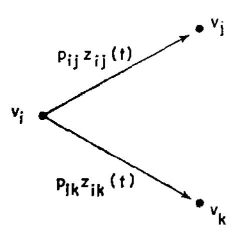

FIG. 4

By way of summary, the process (*) can be geometrically described as a directed flow on a graph or network. The magnitude of the flow at any time depends on the magnitude of the vertex functions at this time, on the normalized and exponentially weighted cross-correlations of the vertex functions at all past times, and on the inputs created by the experimenter.

(E) Deforming a Probabilistic Graph. A closely related geometrical interpretation of (*) can be given if at every time t we think of G as a *probabilistic graph* $G(t)$ with weight $y_{ij}(t)$ assigned to edge e_{ij} [4]. Then (*) becomes a 1-parameter family of probabilistic graphs $\mathscr{G} = \{G(t) : t \geqslant 0\}$, or a continuously differentiable *deformation* of the probabilistic graph $G(0)$. From this perspective, (*) provides a mechanism for continuously deforming one probabilistic graph $G(t_0)$ into another graph $G(t_1)$, $t_1 > t_0$. In particular, when $t_1 = \infty$, we ask for the existence of a limiting graph $G(\infty)$. This question can also be expressed as: when do fluctuations in the transition probabilities $G(t)$ converge to *stationary* transition probabilities $G(\infty)$?

Another probabilistic graph can be constructed from (*) with weight p_{ij} assigned to edge e_{ij}. This graph, called the "coefficient graph" of (*), is the "geometry" of (*) over which the cross-correlated flow passes. When $C \equiv 0$, we shall study the influence of the "geometry" P on the "limiting transition probabilities" $G(\infty)$. P's can be found for which this influence is either negligible or profound.

4. Outstars

In this section, we introduce the simplest example of our prediction theory. This example is characterized by the coefficient matrix

$$P = \begin{pmatrix} 0,\ 1/(n-1),\ 1/(n-1),\ldots,\ 1/(n-1) \\ 0 \end{pmatrix}.$$

The system therefore obeys the equations

$$\dot{x}_1(t) = -\alpha x_1(t) + I_1(t), \tag{10}$$

$$\dot{x}_j(t) = -\alpha x_j(t) + \beta x_1(t-\tau)\, y_{1j}(t) + I_j(t), \qquad j = 2,\ldots, n \tag{11}$$

$$y_{1j}(t) = z_{1j}(t) \left(\sum_{k=2}^{n} z_{1k}(t)\right)^{-1}, \qquad j = 2,\ldots, n, \tag{12}$$

and

$$\dot{z}_{1j}(t) = -u z_{1j}(t) + \beta x_1(t-\tau)\, x_j(t), \qquad j = 2,\ldots, n. \tag{13}$$

*

where all initial data is nonnegative and continuous and $z_{1j}(0) > 0$, $j \neq 1$. All other functions are identically zero, and we let $\sum_{k \neq 1} x_k(0) > 0$ to avoid trivialities. The coefficient graph of (*) is given in Fig. 5. (*) is therefore

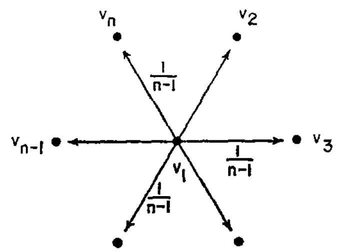

FIG. 5

called an *outstar*. The vertex v_1 is called the *source* of the outstar and each vertex v_j, $j \neq 1$, is called a *sink* of the outstar. The set $B = \{v_j : j \neq 1\}$ of all sinks is called the *border* of the outstar.

Part I studies (*) from a purely mathematical point of view. Part II gives these results a prediction theoretic interpretation. Our mathematical discussion will concern itself with the limiting and oscillatory behavior of (*) as $t \to \infty$ for special choices of the input vector function $C = (I_1, \ldots, I_n)$. These choices will be interpreted in later sections as the presentation to the machine which (*) represents of sequences of predictions to be learned.

The choices of C will be divided into three general cases. In the first case, no inputs reach the border of the outstar at any time. In the second case, inputs do reach this border and continue to do so even at arbitrarily large times. In the third case, inputs reach the border but only for a finite amount of time. All of these cases can be treated by a single method. The success of this method depends on the fact that (*) can be transformed into a more tractable system of equations expressed in terms of new unknown variables. These variables can be classified into two classes. The first class consists of sums over all vertices v_j, $j \neq 1$, in the border (*). These sums are

$$x^{(1)} = \sum_{k \neq 1} x_k, \qquad z^{(1)} = \sum_{k \neq 1} z_{1k}, \qquad \text{and} \qquad I^{(1)} = \sum_{k \neq 1} I_k.$$

The second class consists of three 1-parameter families of probability distributions associated with (*). These are $X = \{X_j : j \neq 1\}$, $y = \{y_{1j} : j \neq 1\}$, and $\theta = \{\theta_j : j \neq 1\}$, where $X_j = x_j/x^{(1)}$, $y_{1j} = z_{ij}/z^{(1)}$, and $\theta_j = I_j/I^{(1)}$. We will find that the sums $x^{(1)}$ and $z^{(1)}$ over the border depend on time only through the known inputs I_1 and $I^{(1)}$. In particular, they are independent of the unknown probabilities X and y. Moreover, (*) can be replaced by a system of equations for the time evolution of the probability distributions $X(t)$ and $y(t)$. The coefficients in these equations depend only on I_1, $I^{(1)}$, and the *known* sums $x^{(1)}$ and $z^{(1)}$. These facts are summarized in the following two lemmas.

LEMMA 1. *The source function x_1 and the sums $x^{(1)}$ and $z^{(1)}$ depend on time only through the known inputs I_1 and $I^{(1)}$.*

PROOF. The assertion is obvious for x_1 by (10). Sum (11) over $j \neq 1$ using the fact that $\sum_{j \neq 1} y_{1j} = 1$. Then

$$\dot{x}^{(1)} = -\alpha x^{(1)} + \beta x_1(t - \tau) + I^{(1)}, \tag{14}$$

and so by (10) the assertion is obvious for $x^{(1)}$. Summing (13) over $j \neq 1$ gives

$$\dot{z}^{(1)} = -u z^{(1)} + \beta x_1(t - \tau)\, x^{(1)}, \tag{15}$$

which gives the assertion for $z^{(1)}$ by (10) and (14).

LEMMA 2. (*) *can be transformed into the following system of equations for the probability distributions y and X.*

$$\dot{X}_j = A_1(y_{1j} - X_j) + B_1(\theta_j - X_j) \tag{16}$$

and

$$\dot{y}_{1j} = C_1(X_j - y_{1j}), \tag{17}$$

where

$$A_1(t) = \frac{\beta x_1(t - \tau)}{x^{(1)}(t)}, \qquad B_1(t) = \frac{I^{(1)}(t)}{x^{(1)}(t)},$$

and

$$C_1(t) = \beta x_1(t - \tau)\frac{x^{(1)}(t)}{z^{(1)}(t)}.$$

PROOF. (16) has the following derivation. Since $X_j = x_j/x^{(1)}$,

$$\dot{X}_j = \frac{1}{x^{(1)}}\left(\dot{x}_j - x_j \frac{\dot{x}^{(1)}}{x^{(1)}}\right).$$

Substituting (11) and (14) into this equation gives

$$\dot{X}_j = \frac{1}{x^{(1)}}\left[-\alpha x_j + \beta x_1(t - \tau)\, y_{1j} + I_j - x_j\left(-\alpha + \frac{\beta x_1(t - \tau) + I^{(1)}}{x^{(1)}}\right)\right]$$
$$= A_1(y_{1j} - X_j) + B_1(\theta_j - X_j).$$

(17) is derived in the following way. Since $y_{1j} = z_{1j}/z^{(1)}$,

$$\dot{y}_{1j} = \frac{1}{z^{(1)}} \left(\dot{z}_{1j} - z_{1j} \frac{\dot{z}^{(1)}}{z^{(1)}} \right).$$

Substituting (13) and (15) into this equation gives

$$\begin{aligned}\dot{y}_{1j} &= \frac{1}{z^{(1)}} \left[- u z_{1j} + \beta x_1(t - \tau)\, x_j - z_{1j} \left(- u + \frac{\beta x_1(t - \tau)\, x^{(1)}}{z^{(1)}} \right) \right] \\ &= C_1(y_{1j} - X_j).\end{aligned}$$

5. Outstars with an Input-Free Border

We use Lemmas 1 and 2 to study the case in which no inputs reach the border of the outstar at any time. Thus $I_j \equiv 0, j \neq 1$, and we say the border of (*) is *input-free*. The main fact needed to carry out our prediction theory in this case is the following.

THEOREM 2. *Let* (*) *be given with arbitrary initial data, an input-free border, and any nonnegative and continuous* I_1. *If* $x_1 \not\equiv 0$, *then* y_{1j} *and* X_j *are monotone in opposite senses and*

$$\lim_{t \to \infty} y_{1j}(t) = \lim_{t \to \infty} X_j(t).$$

If $x_1 \equiv 0$, *then* y_{1j} *and* X_j *are constant.*

PROOF. By (16) and the hypothesis $I^{(1)} \equiv 0$,

$$\dot{X}_j = A_1(y_{1j} - X_j), \tag{16}$$

where $A_1(t) = \beta x_1(t - \tau)/x^{(1)}(t)$ is nonnegative. By (17),

$$\dot{y}_{ij} = C_1(X_j - y_{1j}), \tag{17}$$

where $C_1(t) = \beta x_1(t - \tau)\, x^{(1)}(t)/z^{(1)}(t)$ is nonnegative. From (16) and (17) we draw the following conclusions.

If $x_1 \equiv 0$ then y_{1j} and X_j are constant since $\dot{y}_{1j} = \dot{X}_j \equiv 0$. Suppose that $x_1 \not\equiv 0$. If $X_j(t_0) = y_{1j}(t_0)$, then $X_j(t) = y_{1j}(t) =$ constant for all $t \geqslant t_0$. In particular when $t_0 = 0$, X_j and y_{1j} are constant. Suppose by contrast that $X_j(0) \neq y_{1j}(0)$. By (10), there is a T_0 such that $x_1(t - \tau) = 0$ for $t \in [0, T_0]$ and $x_1(t - \tau) > 0$ for $t > T_0$. If $X_j(0) > y_{1j}(0)$, then $X_j(t)$ and $y_{1j}(t)$ are constant for $t \in [0, T_0]$, but $X_j(t)$ is strictly monotone decreasing and $y_{1j}(t)$ is strictly monotone increasing for all $t \in (T_0, T_1)$, where T_1 is the smallest root, if any, of the equation $X_j(t) = y_{1j}(t)$. If such a T_1 exists, then $X_j(t)$ and $y_{1j}(t)$ are constant for $t \geqslant T_1$. We shall show in the next paragraph that no such T_1 exists. If no such T_1 exists, then $X_j(t)$ decreases monotonically for all $t > T_0$ and $y_{1j}(t)$ increases monotonically for all $t > T_0$. Since X_j and y_{1j} are bounded, the limits $Q_j \equiv \lim_{t\to\infty} X_j(t)$ and $P_{1j} \equiv \lim_{t\to\infty} y_{1j}(t)$ exist. If $X_j(0) < y_{1j}(0)$, the same argument goes through with all inequalities reversed. In all cases, therefore, X_j and y_{1j} are monotone in opposite senses and $| X_j - y_{1j} |$ is monotone nonincreasing.

We now show that T_1 does not exist and that $P_{1j} = Q_j$, $j \neq 1$, if $x_1 \not\equiv 0$. Subtracting (16) from (17) gives

$$(y_{1j} - X_j)^{\boldsymbol{\cdot}} = - D_1(y_{1j} - X_j), \tag{18}$$

where

$$D_1 \equiv A_1 + C_1 = \beta x_1(t - \tau) \left(\frac{1}{x^{(1)}} + \frac{x^{(1)}}{z^{(1)}}\right).$$

Integrating (18), we find

$$y_{1j}(t) - X_j(t) = (y_{1j}(0) - X_j(0))\, \Omega_1(t), \tag{19}$$

where

$$\Omega_1(t) = \exp\left[-\beta \int_0^t x_1(v - \tau)\left(\frac{1}{x^{(1)}(v)} + \frac{x^{(1)}(v)}{z^{(1)}(v)}\right) dv\right].$$

To show that T_1 does not exist, note that $\Omega_1(t) > 0$, $t \geqslant 0$. Thus $y_{1j}(0) \neq X_j(0)$ implies $y_{1j}(t) \neq X_j(t)$. To show that $P_{1j} = Q_j$, we must show that $\lim_{t\to\infty} \Omega_1(t) = 0$, or that

$$\lim_{t\to\infty} \int_0^t x_1(v - \tau)\left(\frac{1}{x^{(1)}(v)} + \frac{x^{(1)}(v)}{z^{(1)}(v)}\right) dv = \infty.$$

Since $x^{(1)}/z^{(1)}$ is positive, it suffices to show that

$$\lim_{t\to\infty} \int_0^t \frac{x_1(v - \tau)}{x^{(1)}(v)}\, dv = \infty.$$

For $t \geqslant 0$,

$$\begin{aligned}
\beta \int_0^t \frac{x_1(v-\tau)}{x^{(1)}(v)}\, dv &= \int_0^t \frac{\beta x_1(v-\tau)}{e^{-\alpha v}\left[x^{(1)}(0) + \beta \int_0^v x_1(w-\tau)\, e^{\alpha w}\, dw\right]}\, dv \\
&= \int_0^t \frac{d}{dv} \log\left(x^{(1)}(0) + \beta \int_0^v x_1(w-\tau)\, e^{\alpha w}\, dw\right) dv \\
&= \log\left(1 + \frac{\beta}{x^{(1)}(0)} \int_0^t x_1(v-\tau)\, e^{\alpha v}\, dv\right).
\end{aligned}$$

By (10) and the nonnegativity of I_1, $\dot{x}_1 \geqslant -\alpha x_1$. Thus

$$\int_0^t \frac{x_1(v-\tau)}{x^{(1)}(v)}\, dv$$

diverges at a logarithmic rate as $t \to \infty$, and $P_{1j} = Q_j$, $j \neq 1$.

Theorem 2 is summarized in Fig. 6. It shows that the limits $\lim_{t\to\infty} y_{ij}(t)$ and $\lim_{t\to\infty} X_j$ do not vary continuously as a function of the initial data $x_1(v)$, $v \in [-\tau, 0]$. This theorem is picturesquely called the "speck of dust"

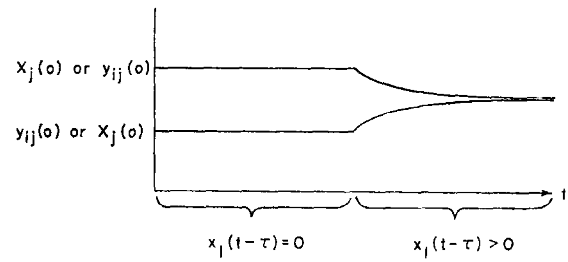

FIG. 6

theorem, because it describes an alternative which depends on whether or not the source function x_1 is identically zero. Since I_1 is nonnegative, the positivity of $x_1(t_0)$ for any t_0 implies the positivity of $x_1(t)$ for all $t \geqslant t_0$. Thus if the initial data of x_1 is identically zero, then $x_1(t)$ remains zero until a positive value of $I_1(t)$, no matter how small—that is, a "speck of dust"—reaches the source v_1. Thereafter $x_1(t)$ remains positive at all times.

By Theorem 2, if $X_j(0) = y_{1j}(0)$ then $X_j(t) = y_{1j}(t) =$ constant for all $t \geqslant 0$ and any choice of I_1. This means in particular that arbitrary probability distributions can arise as limits $\lim_{t\to\infty} X_j(t) = \lim_{t\to\infty} y_{1j}(t)$, $j \neq 1$. The coefficient matrix P of an outstar thus does not uniquely determine the limiting distributions when the border of the outstar is input-free. That is, the "geometry" P has little effect on the "limiting transition probabilities" $G(\infty)$.

Theorem 2 contains all the information required for our simplest prediction theoretic needs, and we therefore recommend that the reader interested primarily in the prediction theory go on immediately to the next section.

More information is available concerning an outstar with input-free border than is contained in Theorem 2, because (*) can be explicitly integrated in this case to give precise information about the relative rates at which the probability distributions associated with different vertices and edges approach their limits as $t \to \infty$. A brute force integration of (*) by an exponential change of variable seems to indicate that each $x_j(t)$ and $z_{1j}(t)$ depends on all its past values $x_j(v)$ and $z_{1j}(v)$, $0 \leqslant v \leqslant t$, in a complicated way. This is, however, not so.

THEOREM 3. *x_j and z_{1j} obey equations of the form*

$$x_j(t) = X_j(0)\, a(t) + (y_{1j}(0) - X_j(0))\, b(t) \tag{20}$$

and

$$z_{1j}(t) = X_j(0)\, c(t) + (y_{1j}(0) - X_j(0))\, d(t), \tag{21}$$

where a, b, c, and d are nonnegative and continuous functions that depend only on I_1 and the initial data $x_1(0)$, $x^{(1)}(0)$, and $z^{(1)}(0)$.

PROOF. (19) can be written as

$$\frac{z_{1j}}{z^{(1)}} - \frac{x_j}{x^{(1)}} = E_j\,, \tag{22}$$

where $E_j = k_j\Omega_1$ and $k_j = y_{1j}(0) - X_j(0)$. E_j is a known function, since by Lemma 1, Ω_1 can be written as an explicit integral which depends on time only through I_1, and on the initial data only through $x_1(0)$, $x^{(1)}(0)$, and $z^{(1)}(0)$. By (22),

$$z_{1j} = Ux_j + V_j\,, \tag{23}$$

where $U = z^{(1)}/x^{(1)}$ and $V_j = z^{(1)}E_j$ are known functions. Differentiating (23) we find

$$\dot{z}_{1j} = \dot{U}x_j + U\dot{x}_j + \dot{V}_j\,.$$

(13) provides another expression for $\dot{z}_{1j}$. Identifying these two expressions and rearranging terms gives

$$U\dot{x}_j + (uU + \dot{U} - \beta x_1(t-\tau))\, x_j + (uV_j + \dot{V}_j) = 0.$$

Since $U > 0$, division by U is permissible and we find

$$\dot{x}_j + Px_j + Q_j = 0 \tag{24}$$

where

$$P = u + \frac{d}{dt}\log U - \frac{\beta x_1(t-\tau)}{U}\,,$$

$$Q_j = k_j Q,$$

and

$$Q = \left[u + \frac{d}{dt}\log z^{(1)} - \beta x_1(t-\tau)\left(\frac{x^{(1)}}{z^{(1)}} + \frac{1}{x^{(1)}}\right)\right] x^{(1)}\Omega_1\,.$$

The coefficients P and Q appearing in (24) are known functions which depend on time only through I_1, (24) can therefore be integrated. We find

$$x_j(t) = x_j(0)\exp\left[-\int_0^t P(u)\,du\right] - k_j\int_0^t Q(v)\exp\left[-\int_v^t P(u)\,du\right] dv. \tag{25}$$

(25) can be simplified by noticing that

$$\underset{\sim}{Q} = \left(u + \frac{\dot{z}^{(1)} - \beta x_1(t-\tau)\, x^{(1)}}{z^{(1)}} - \frac{\beta x_1(t-\tau)}{x^{(1)}}\right) x^{(1)} \Omega_1$$

$$= \left(u + \frac{-uz^{(1)}}{z^{(1)}} - \frac{\beta x_1(t-\tau)}{x^{(1)}}\right) x^{(1)} \Omega_1$$

$$= -\beta x_1(t-\tau)\, \Omega_1 ,$$

and that

$$P = u + \frac{d}{dt} \log \frac{z^{(1)}}{x^{(1)}} - \frac{\beta x_1(t-\tau)\, x^{(1)}}{z^{(1)}}$$

$$= u + \frac{\dot{z}^{(1)} - \beta x_1(t-\tau)\, x^{(1)}}{z^{(1)}} - \frac{d}{dt} \log x^{(1)}$$

$$= -\frac{d}{dt} \log x^{(1)}.$$

Hence

$$\exp\left[-\int_0^t P(u)\, du\right] = \exp\left[\log \frac{x^{(1)}(t)}{x^{(1)}(0)}\right] = \frac{x^{(1)}(t)}{x^{(1)}(0)}$$

and

$$\exp\left[-\int_v^t P(u)\, du\right] = \frac{x^{(1)}(t)}{x^{(1)}(v)} .$$

Substituting these simplified expressions into (25) gives

$$x_j(t) = x_j(0) \frac{x^{(1)}(t)}{x^{(1)}(0)} + \beta k_j \int_0^t x_1(v-\tau)\, \Omega_1(v) \frac{x^{(1)}(t)}{x^{(1)}(v)}\, dv. \tag{26}$$

Let $a(t) = x^{(1)}(t)$ and

$$b(t) = \beta x^{(1)}(t) \int_0^t \frac{x_1(v-\tau)\, \Omega_1(v)}{x^{(1)}(v)}\, dv$$

in (26). $a(t)$ depends only on I_1 and on the initial data $x^{(1)}(0)$. $b(t)$ depends only on I_1 and the initial data $x_1(0)$, $x^{(1)}(0)$, and $z^{(1)}(0)$. This completes the proof of (20).

We now use (20) to derive an equation for z_{1j}. By (20) and (23),

$$z_{1j} = Ux_j + V_j$$

$$= Ux^{(1)} \left(X_j(0) + \beta k_j \int_0^t \frac{x_1(v-\tau)\, \Omega_1(v)}{x^{(1)}(v)}\, dv\right) + k_j z^{(1)} \Omega_1 .$$

Since $U = z^{(1)}/x^{(1)}$,

$$z_{1j} = z^{(1)} \left[X_j(0) + k_j \left(\Omega_1 + \beta \int_0^t \frac{x_1(v-\tau)\, \Omega_1(v)}{x^{(1)}(v)}\, dv\right)\right]. \tag{27}$$

Letting

$$c(t) = z^{(1)} \qquad \text{and} \qquad d(t) = z^{(1)}(t) \left[\Omega_1(t) + \beta \int_0^t \frac{x_1(v-\tau)\, \Omega_1(v)}{x^{(1)}(v)}\, dv\right]$$

in (27) completes the proof of (21) once we observe that $c(t)$ and $d(t)$ depend only on I_1 and on the initial data $x_1(0)$, $x^{(1)}(0)$, and $z^{(1)}(0)$.

Theorem 3 can be used to show that the probability distributions at different vertices and edges approach their limits at the same rate, except for a multiplicative factor that depends on their initial data, as we prove in the next corollary. We already know that $y_{1i}(0) = X_1(0)$ implies

$$y_{1i}(t) = X_i(t) = \text{constant},$$

$t \geqslant 0$, and that $y_{1i}(t)$ and $X_i(t)$ are constant whenever $x_1(t-\tau) = 0$. We consider therefore only the remaining case in which $y_{1i}(0) \neq X_i(0)$ and $x_1(t-\tau) > 0$.

COROLLARY 2. *Suppose $y_{1i}(0) \neq X_i(0)$, $y_{1j}(0) \neq X_j(0)$, and $x_1(t_0 - \tau) > 0$. Then*

$$\frac{y_{1i}(t) - X_i(0)}{y_{1j}(t) - X_j(0)} = \frac{X_i(t) - X_i(0)}{X_j(t) - X_j(0)} = \frac{y_{1i}(t) - X_i(t)}{y_{1j}(t) - X_j(t)} = \text{constant}, \qquad t \geqslant t_0.$$

PROOF. The proof of Theorem 3 shows that $a = x^{(1)}$ and $c = z^{(1)}$. Thus (20) and (21) can be written as

$$X_j(t) - X_j(0) = (y_{1j}(0) - X_j(0)) \frac{b(t)}{a(t)},$$

and

$$y_{1j}(t) - X_j(0) = (y_{1j}(0) - X_j(0)) \frac{d(t)}{c(t)},$$

where $b(t)/a(t) > 0$ and $d(t)/c(t) > 0$ for $t \geqslant t_0$ since $x_1(t-\tau) > 0$. Thus

$$\frac{X_i(t) - X_i(0)}{X_j(t) - X_j(0)} = \frac{y_{1i}(0) - X_i(0)}{y_{1j}(0) - X_j(0)} = \frac{y_{1i}(t) - X_i(0)}{y_{1j}(t) - X_j(0)}$$

for $t \geqslant t_0$. By (19) we also know that

$$\frac{y_{1i}(t) - X_i(t)}{y_{1j}(t) - X_j(t)} = \frac{y_{1i}(0) - X_i(0)}{y_{1j}(0) - X_j(0)}.$$

Combining these equalities completes the proof.

6. Outstars Whose Border Never Becomes Input-Free

In the preceding section, we found that any probability distribution $y_{1j}(t) = X_j(t)$, $j \neq 1$, remains constant for all $t \geqslant 0$ when the outstar's border is input-free. This fact provides an affirmative answer to the following question. Given any probability distribution $\{\theta_j, j \neq 1\}$, does there exist an input function $C = (I_1, ..., I_n)$ for which $X_j(0) = y_{1j}(0) = \theta_j$ and

$$\lim_{t\to\infty} X_j(t) = \lim_{t\to\infty} y_{1j}(t) = \theta_j ?$$

Any nonnegative C with $I_j = 0$, $j \neq 1$, accomplishes this goal. A natural generalization of this question is the following question. Given any three probability distributions $\theta_j^{(1)}$, $\theta_j^{(2)}$, and $\theta_j^{(3)}$, $j \neq 1$, does there exist an input vector function C for which

$$X_j(0) = \theta_j^{(1)}, \qquad y_{1j}(0) = \theta_j^{(2)},$$

and

$$\lim_{t\to\infty} X_j(t) = \lim_{t\to\infty} y_{1j}(t) = \theta_j^{(3)} ?$$

We now answer this question in the affirmative and provide a considerable amount of supplementary information concerning the manner in which the probability distributions X_j and y_{1j} approach their limits. We do this in the following theorem.

Theorem 4. *Suppose $\alpha > 0$ and $u > 0$. Let the inputs to the border of an outstar have the form $I_j(t) = \theta_j I(t)$, $j \neq 1$, where $\{\theta_j : j \neq 1\}$ is a fixed, but arbitrary, probability distribution, and $I_1(t)$ and $I(t)$ are arbitrary nonnegative and continuous functions. Then the functions $f_j(t) = y_{1j}(t) - X_j(t)$, $g_j(t) = X_j(t) - \theta_j$, and $\dot{y}_{1j}(t)$ change sign at most once, and not at all if $f_j(0)\, g_j(0) \geqslant 0$. Moreover $f_j(0)\, g_j(0) > 0$ implies $f_j(t)\, g_j(t) > 0$ for all $t \geqslant 0$. Suppose furthermore that $I_1(t)$ and $I(t)$ are bounded functions, and that there exist two positive constants c and T_0 such that*

$$\int_0^t e^{-\alpha(t-v)} I_1(v)\, dv \geqslant c$$

and

$$\int_0^t e^{-\alpha(t-v)} I(v)\, dv \geqslant c$$

for $t \geqslant T_0$. Then

$$\lim_{t\to\infty} X_j(t) = \lim_{t\to\infty} y_{1j}(t) = \theta_j, \qquad j \neq 1.$$

PROOF. The proof is divided into three steps. In step (I) we prove that f_j, g_j, and $\dot{y}_{1j}$ change sign at most once and, thus, that $\lim_{t\to\infty} y_{1j}(t)$ exists, $j \neq 1$. In step (II), the existence of these limits along with estimates of $A_1(t)$, $B_1(t)$, $C_1(t)$, and $\dot{C}_1(t)$ for large t are used to show that the limits $\lim_{t\to\infty} X_j(t)$ exist and equal the limits $\lim_{t\to\infty} y_{1j}(t)$. In step (III), the common value of these limits is shown to be θ_j by estimating $A_1(t)$, $B_1(t)$, $\dot{A}_1(t)$, and $\dot{B}_1(t)$ for large t.

(I) Subtracting (16) from (17) gives

$$\dot{f}_j = - D_1 f_j + B_1 g_j , \tag{28}$$

where $D_1 = A_1 + C_1$. Since $(X_j - \theta_j)^{\cdot} = \dot{X}_j$, (16) can be written as

$$\dot{g}_j = - B_1 g_j + A_1 f_j . \tag{29}$$

Equations (28) and (29) are special cases of the following simple but basic lemma.

LEMMA 3. *Let the functions f and g satisfy the differential equations*

$$\dot{f} = af + bg$$

$$\dot{g} = cf + dg,$$

where a, b, c, and d are continuous functions and the off-diagonal coefficients b and c are nonnegative. Then f and g change sign at most once and not at all if $f(0)\,g(0) \geqslant 0$. Moreover $f(0)\,g(0) > 0$ implies $f(t)\,g(t) > 0$ for all $t \geqslant 0$.

Lemma 3 can be geometrically visualized by Fig. 7 which shows the (f, g) plane. The direction of the arrows indicates the path of the (f, g) point through time.

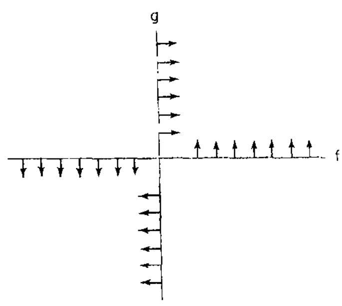

FIG. 7

PROOF. Clearly $(fg)^{\cdot} = (a + d) fg + bg^2 + cf^2 \geqslant (a + d) fg$ by the nonnegativity of b and c. Thus for any $t_0 \geqslant 0$, $(fg)(t_0) \geqslant 0$ implies $(fg)(t) \geqslant 0$ for all $t \geqslant t_0$. More can be said. Let $(fg)(t_0) > 0$, where $f(t_0) > 0$ (say), and let $t = t_1 > t_0$ be the first zero of f, or g, or both (of f, say). Then f and g are both nonnegative in $[t_0, t_1]$, so $\dot{f} \geqslant af$ in $[t_0, t_1]$, and

$$0 = f(t_1) \geqslant e^{a(t-t_0)} f(t_0) > 0,$$

which is a contradiction. Thus $(fg)(t) > 0$ for all $t \geqslant t_0$ if $(fg)(t_0) > 0$.

Only the case $(fg)(0) < 0$ remains, where $f(0) < 0$ (say). Then either $f(t) < 0 < g(t)$ for all $t \geqslant 0$, or there is a first $t_1 > 0$ when f, or g, or both have a zero. If such a t_1 exists, we are in a previous case, so that f and g change sign at most once.

Lemma 3 can be directly applied to (28) and (29) by letting $f = f_j$, $g = g_j$, $a = -D_1$, $b = B_1$, $c = A_1$, and $d = -B_1$. We conclude that f_j and g_j change sign at most once and not at all if $(f_j g_j)(0) \geqslant 0$. Moreover, $(f_j g_j)(0) > 0$ implies $(f_j g_j)(t) > 0$ for all $t \geqslant 0$.

By (17),

$$\dot{y}_{1j} = -C_1 f_j .$$

Since C_1 is nonnegative, $\dot{y}_{1j}$ also changes sign at most once and not at all if $(f_j g_j)(0) \geqslant 0$. In particular, there exists a T_1 such that $y_{1j}(t)$ is a monotonic function for $t \geqslant T_1$. y_{1j} is also bounded and continuous. Thus $\lim_{t\to\infty} y_{1j}(t)$ exists for all $j \neq 1$.

(II) Using the facts proved in (I), we now show that the limits $\lim_{t\to\infty} X_j(t)$ exist. The first step in this proof is to establish various estimates for the coefficients A_1, B_1, and C_1 which appear in (16) and (17). The purpose of these estimates is to show that $\ddot{y}_{1j}(t)$ is bounded for sufficiently large t. This fact, in turn, will be needed to prove that $\lim_{t\to\infty} \dot{y}_{1j}(t) = 0$, from which it will follow with the help of the estimates that $\lim_{t\to\infty} X_j(t)$ exists and equals $\lim_{t\to\infty} y_{1j}(t)$.

The estimates needed for A_1, B_1, and C_1 are the following. We shall find positive constants λ_i, $i = 1, 2, 3, 4, 5$, and a time T_2 such that for $t \geqslant T_2$, the inequalities $\lambda_1 \leqslant C_1(t) \leqslant \lambda_2$, $A_1(t) \leqslant \lambda_3$, $B_1(t) \leqslant \lambda_4$, and $|\dot{C}_1(t)| \leqslant \lambda_5$ hold. To establish these estimates, we make comparable estimates on the functions x_1, $x^{(1)}$, and $z^{(1)}$ from which A_1, B_1, and C_1 are constructed. Firstly we establish lower bounds for these functions for large t.

By hypothesis, there exist positive constants c and T_0 such that

$$\int_0^t e^{-\alpha(t-v)} I_1(v)\, dv \geqslant c \qquad \text{for} \qquad t \geqslant T_0 .$$

Thus by integrating (10) we find

$$x_1(t) \geqslant e^{-\alpha t} x_1(0) + c \geqslant c \qquad \text{for} \qquad t \geqslant T_0 .$$

Substituting this inequality into the integrated form of (14) gives for $t \geqslant 2(T_0 + \tau)$,

$$x^{(1)}(t) \geqslant e^{-\alpha t}\left(x^{(1)}(0) + \int_0^{T_0+\tau} e^{\alpha v}(\beta x_1(v-\tau) + I^{(1)}(v))\,dv + \beta c \int_{T_0+\tau}^{\tau} e^{\alpha v}\,dv\right)$$

$$\geqslant \frac{\beta c}{\alpha}(1 - e^{-\alpha(T_0+\tau)}) \equiv d > 0.$$

Substituting these inequalities into the integrated form of (15) we find for $t \geqslant 3(T_0 + \tau)$ that

$$z^{(1)}(t) \geqslant e^{-ut}\left(z^{(1)}(0) + \int_0^{2(T_0+\tau)} e^{uv} x_1(v-\tau)\, x^{(1)}(v)\,dv + \beta cd \int_{2(T_0+\tau)}^{t} e^{uv}\,dv\right)$$

$$\geqslant \frac{\beta cd}{u}(1 - e^{-u(T_0+\tau)}) \equiv e > 0.$$

Upper bounds for x_1, $x^{(1)}$, and $z^{(1)}$ follow by the boundedness of I_1 and $I^{(1)}$. Letting $M = \sup\{I_1(t) : t \geqslant 0\}$ and $M^{(1)} = \sup\{I(t) : t \geqslant 0\}$, we readily find that

$$x_1(t) \leqslant x_1(0) + \frac{M_1}{\alpha} \equiv M < \infty,$$

$$x^{(1)}(t) \leqslant x^{(1)}(0) + \frac{\beta}{\alpha}(M + M^{(1)}) \equiv N < \infty,$$

and

$$z^{(1)}(t) \leqslant z^{(1)}(0) + \frac{\beta MN}{u} \equiv R < \infty.$$

Let $T_2 = 3(T_0 + \tau)$. Then the following definitions of the λ_i, $i = 1, 2, 3, 4$ obviously suffice for $t \geqslant T_2 : \lambda_1 = \beta cd/R$, $\lambda_2 = \beta MN/e$, $\lambda_3 = \beta M/d$, and $\lambda_4 = M^{(1)}/d$.

$\dot{x}_1$, $\dot{x}^{(1)}$, and $\dot{z}^{(1)}$ can also be shown to be bounded by simple estimates of the above kind. Using these estimates along with those derived above readily shows that there exists a $\lambda_5 < \infty$ such that $|\dot{C}_1(t)| \leqslant \lambda_5$ for all $t \geqslant T_2$.

These various estimates on the functions A_1, B_1, C_1, and $\dot{C}_1$ suffice to show that $|\ddot{y}_{ij}(t)|$ is bounded for $t \geqslant T_2$, since by (16) and (17),

$$\begin{aligned}|\ddot{y}_{1j}| &= |\dot{C}_1(X_j - y_{1j}) + C_1(\dot{X}_j - \dot{y}_{1j})| \\ &\leqslant 2|\dot{C}_1| + |C_1|\,[(A_1 + C_1)\,|y_{1j} - X_j| + B_1\,|\theta_j - X_j|) \\ &\leqslant 2(\lambda_5 + \lambda_2(\lambda_2 + \lambda_3 + \lambda_4)) < \infty,\end{aligned}$$

for $t \geqslant T_2$.

To show that $\lim_{t\to\infty} \dot{y}_{1j}(t) = 0$, we need the following simple lemma.

LEMMA 4. *Suppose* $f(t) \to \alpha < \infty$ *as* $t \to \infty$ *and* $\ddot{f}$ *is bounded. Then* $\dot{f}(t) \to 0$ *as* $t \to \infty$.

PROOF. Suppose not. Then for some $\epsilon > 0$, there exists a sequence $\{t_n\}$ with $\lim_{n\to\infty} t_n = \infty$ such that $|\dot{f}(t_n)| \geqslant \epsilon$ for all n. We can suppose $\dot{f}(t_n) \geqslant \epsilon$ for all n without loss of generality. Since $\ddot{f}$ is bounded, there exists a δ such that $\dot{f} \geqslant (\epsilon/2)$ on infinitely many nonoverlapping intervals $I_n = [u_n, u_n + \delta]$ of length δ, where $\lim_{n\to\infty} u_n = \infty$. Thus $f(u_n + \delta) - f(u_n) \geqslant (\epsilon\delta/2)$ for all n, and $f \not\to \alpha < \infty$ as $t \to \infty$, which is a contradiction.

Replacing f by y_{1j} in Lemma 4 immediately shows that $\lim_{t\to\infty} \dot{y}_{1j}(t) = 0$.

We can now show the existence of $\lim_{t\to\infty} X_j(t)$. By (I) we can assume that $\dot{y}_{1j}(t) \geqslant 0$ and thus that $(X_j - y_{1j})(t) \geqslant 0$ for $t \geqslant T_0$ without loss of generality. Also $C_1(t) \geqslant \lambda_1 > 0$ for $t \geqslant T_2$. Thus by (17).

$$\dot{y}_{1j} = C_1(X_j - y_{1j}) \geqslant \lambda_1(X_j - y_{1j}) \geqslant 0$$

for $t \geqslant \max(T_0, T_2)$. Since $\lim_{t\to\infty} \dot{y}_{1j}(t) = 0$, it follows immediately that $\lim_{t\to\infty}(X_j(t) - y_{1j}(t)) = 0$. We also know that $\lim_{t\to\infty} y_{1j}(t)$ exists, by (I). Thus $Q_j \equiv \lim_{t\to\infty} X_j(t)$ exists and equals $\lim_{t\to\infty} y_{1j}(t)$.

(III) Letting $X_j^{(\theta)} = X_j - \theta_j$ and $F_{1j} = A_1(y_{1j} - X_j)$, we integrate (16) to find for every $T \geqslant 0$ and $t \geqslant T$ that

$$X_j^{(\theta)}(t) = e^{-\int_T^t B_1 dv} \left[X_j^{(\theta)}(T) + \int_T^t e^{\int_T^v B_1 dw} F_{1j}\, dv\right].$$

To establish the equality $Q_j = \theta_j$, we must prove $\lim_{t\to\infty} X_j^{(\theta)}(t) = 0$. To do this, we will find positive constants μ and ν such that for $t \geqslant T_0$,

$$\mu + \alpha t \leqslant \int_0^t B_1 dv \leqslant \nu + \alpha t.$$

From this follows that for every fixed $T_3 \geqslant T_0$,

$$\lim_{t\to\infty} e^{-\int_{T_0}^t B_1 dv} \left[X_j^{(\theta)}(T_0) + \int_{T_0}^{T_3} e^{\int_{T_0}^v B_1 dw} F_{1j}\, dv\right] = 0.$$

It remains only to show that there exists a T_3 such that

$$\lim_{t\to\infty} e^{-\int_{T_0}^t B_1 dv} \int_{T_3}^t e^{\int_{T_0}^v B_1 dw} F_{1j}\, dv = 0.$$

Since A_1 is bounded and

$$\lim_{t\to\infty} y_{1j}(t) = \lim_{t\to\infty} X_j(t) \leqslant 1,$$

$$\lim_{t\to\infty} F_{1j}(t) = 0.$$

Thus for every $\epsilon > 0$ there is a $T_3 \geqslant T_0$ such that $|F_{1j}(t)| \leqslant \epsilon e^{\mu-\nu}$ for $t \geqslant T_3$, and we find

$$\left| e^{-\int_{T_0}^{t} B_1 dv} \int_{T_3}^{t} e^{\int_{T_0}^{v} B_1 dw} F_{1j}\, dv \right| \leqslant \epsilon.$$

It remains only to estimate $\int_0^t B_1\, dv$. Since $B_1 = I/x^{(1)}$, we consider $x^{(1)}$. We know that

$$x^{(1)}(t) = e^{-\alpha t}\left[x^{(1)}(0) + \int_0^t e^{\alpha v}(\beta x_1(v-\tau) + I(v))\, dv\right],$$

where

$$x_1(v-\tau) = e^{-\alpha v}\left[x_1(-\tau) + e^{-\alpha\tau}\int_{-\tau}^{v-\tau} e^{\alpha w} I_1(w)\, dw\right].$$

Thus

$$e^{\alpha t}x^{(1)}(t) = x^{(1)}(0) + \int_0^t e^{\alpha v} I(v)\, dv$$
$$+ \beta x_1(-\tau)\, t + \beta e^{-\alpha\tau}\int_{-\tau}^{t-\tau} dv \int_{-\tau}^{v} e^{\alpha w} I_1(w)\, dw.$$

We estimate $e^{\alpha t}x^{(1)}(t)$ term-by-term. Since I_1 is bounded, there exists a constant C such that

$$\int_0^t e^{\alpha w} I_1(w)\, dw \leqslant C e^{\alpha t}.$$

Thus

$$\int_{-\tau}^{t-\tau} dv \int_{-\tau}^{v} e^{\alpha w} I_1(w)\, dw \leqslant C\int_{-\tau}^{t-\tau} e^{\alpha v}\, dv = \frac{Ce^{-\alpha\tau}}{\alpha}(e^{\alpha t} - 1),$$

and we can find a constant k such that

$$\int_{-\tau}^{t-\tau} dv \int_{-\tau}^{v} e^{\alpha w} I_1(w)\, dw \leqslant k e^{\alpha t}.$$

Thus for $t \geqslant T_0$,

$$\int_{-\tau}^{t-\tau} dv \int_{-\tau}^{v} e^{\alpha w} I_1(w)\, dw \leqslant \frac{k}{c}\int_0^t e^{\alpha v} I(v)\, dv.$$

Also there is certainly a K such that for $t \geqslant T_0$,

$$t \leqslant Ke^{\alpha t} \leqslant \frac{K}{c} \int_0^t e^{\alpha v} I(v)\, dv.$$

From these estimates, we readily find constants ω_1 and ω_2 such that

$$e^{\alpha t} x^{(1)}(t) \leqslant \omega_1 + \omega_2 \int_0^t e^{\alpha v} I(v)\, dv$$

for $t \geqslant 0$. By the nonnegativity of I_1 and $x_1(-\tau)$, it is also immediate that

$$e^{\alpha t} x^{(1)}(t) \geqslant x^{(1)}(0) + \int_0^t e^{\alpha v} I(v)\, dv.$$

Thus

$$\frac{Ie^{\alpha t}}{\omega_1 + \omega_2 \int_0^t e^{\alpha v} I(v)\, dv} \leqslant B_1(t) \leqslant \frac{Ie^{\alpha t}}{x^{(1)}(0) + \int_0^t e^{\alpha v} I(v)\, dv},$$

or

$$\frac{1}{\omega_2} \frac{d}{dt} \log \left(\omega_1 + \omega_2 \int_0^t e^{\alpha v} I(v)\, dv\right) \leqslant B_1(t) \leqslant \frac{d}{dt} \log \left(x^{(1)}(0) + \int_0^t e^{\alpha v} I(v)\, dv\right)$$

Integrating, we find

$$\frac{1}{\omega_2} \log \left(1 + \frac{\omega_2}{\omega_1} \int_0^t e^{\alpha v} I(v)\, dv\right) \leqslant \int_0^t B_1\, dv \leqslant \log \left(1 + \frac{1}{x^{(1)}(0)} \int_0^t e^{\alpha v} I(v)\, dv\right).$$

Thus for $t \geqslant T_0$,

$$\frac{1}{\omega_2} \log \left(1 + \frac{\omega_2 c}{\omega_1} e^{\alpha t}\right) \leqslant \int_0^t B_1\, dv \leqslant \log \left(1 + \frac{C}{x^{(1)}(0)} e^{\alpha t}\right),$$

from which the existence of constants μ and ν such that

$$\mu + \alpha t \leqslant \int_0^t B_1\, dv \leqslant \nu + \alpha t$$

for $t \geqslant T_0$ readily follows, and the proof is complete.

The following corollary to Theorem 2 will have a useful prediction theoretic interpretation. This corollary discusses the effect of choosing $I_1(t)$ and $I(t)$ to be periodic successions of "input pulses," where an *input pulse* $J(t)$ is a nonnegative and continuous function which is positive in a finite interval, COROLLARY 3. *Let the functions $I_1(t)$ and $I(t)$ of Theorem 2 be defined as follows.*

$$I_1(t) = \sum_{k=0}^{\infty} J_1(t - k(w + W)) \qquad \textit{and} \qquad I(t) = \sum_{k=0}^{\infty} J_2(t - w - k(w + W)),$$

where $J_i(t)$ is an arbitrary input pulse which is positive in an interval $(0, \lambda_i)$, $\lambda_i > 0$, $i = 1, 2$, and w and W are nonnegative numbers whose sum is positive. Then all the conclusions of Theorem 2 hold.

PROOF. It is obvious that I_1 and I are nonnegative, continuous, and bounded functions. It remains only to find positive constants c and T_0 such that, for example,

$$\Phi(t) \equiv \int_0^t e^{-\alpha(t-v)} I_1(v)\, dv \geqslant c \qquad \text{for} \qquad t \geqslant T_0 .$$

Writing $p = w + W$, let

$$F(t) = \int_{t-p}^{t} e^{-\alpha(t-v)} I_1(v)\, dv, \qquad t \geqslant p.$$

Then for any $n \geqslant 1$ and $t \in [np, (n + 1)p)$,

$$\Phi(t) \geqslant F(t) + e^{-\alpha p} F(t - p) + \cdots + e^{-\alpha(n-1)p} F(t - (n - 1)p).$$

Clearly $F(t) \geqslant F(t - p)$ for all $t \geqslant 2p$, since $I_1(t) \geqslant I_1(t - p)$. Thus

$$\begin{aligned} \Phi(t) &\geqslant (1 + e^{-\alpha p} + \cdots + e^{-\alpha(n-1)p}) F(t - (n - 1)p) \\ &\geqslant F(t - (n - 1)p), \end{aligned}$$

for any $t \in [np, (n + 1)p)$. Since $F_n(t) \equiv F(t - (n - 1)p)$ is a positive and continuous function of $t \in [p, 2p]$, letting $c = \inf\{F_n(t) : t \in [p, 2p]\}$ (> 0) and $T_0 = p$ completes the proof.

7. OUTSTARS WHOSE BORDER EVENTUALLY BECOMES INPUT-FREE

In the previous section, we considered outstars subjected to inputs to border vertices of the form $I_j(t) = \theta_j I(t)$ which take on positive values at arbitrarily large times. Such an outstar is said to be a $G^{(\infty)}$ outstar, and we affix the superscript "(∞)" to each of its functions. For example, we write I_1 as $I_1^{(\infty)}$, x_j as $x_j^{(\infty)}$, and so on.

The border of $G^{(\infty)}$ never becomes input-free. In this section, we consider outstars whose border does eventually become input free. Given any outstar of type $G^{(\infty)}$, we shall construct an infinite sequence of outstars $G^{(1)}, G^{(2)}, \ldots,$

$G^{(N)}, \ldots$, each with the same initial data as $G^{(\infty)}$, and each with a border which eventually becomes input-free. We shall then study the limiting and oscillatory behavior of the functions of $G^{(N)}$ as N and t are permitted to become large by comparing these functions with those of $G^{(\infty)}$ and of outstars with input-free borders. $G^{(N)}$ is defined in terms of a given outstar of type $G^{(\infty)}$ and two given functions $U_1(N)$ and $U(N)$ of $N \geqslant 1$ which are positive and monotone increasing with $\lim_{N\to\infty} U_1(N) = \lim_{N\to\infty} U(N) = \infty$. $G^{(N)}$ has the same initial data as $G^{(\infty)}$. The input functions of $G^{(N)}$ are

$$I_1^{(N)}(t) = I_1(t)\,\chi(t - U_1(N)) \qquad \text{and} \qquad I^{(N)}(t) = I(t)\,\chi(t - U(N)),$$

where $\chi(w) = 1 - \theta(w)$. $G^{(N)}$ is called an *N-truncation* of $G^{(\infty)}$ because its functions agree with those of $G^{(\infty)}$ in the interval $[0, \min(U_1(N), U(N))]$. The border of $G^{(N)}$ is also eventually input-free since no inputs reach the border in the interval $[U(N), \infty]$. We denote the X_j and y_{1j} functions of $G^{(N)}$ by $X_j^{(N)}$ and $y_{1j}^{(N)}$, respectively. The following theorem holds for these functions.

(7A) The Probability Distributions of an Outstar $G^{(N)}$ Remain Essentially Fixed for Large Times

THEOREM 5. *Let $G^{(1)}, G^{(2)}, \ldots, G^{(N)}, \ldots$ be any sequence of N-truncations of any outstar of type $G^{(\infty)}$. Then*

(I) *for every $N = 1, 2, \ldots$, the limits $\lim_{t\to\infty} X_j^{(N)}(t)$ and $\lim_{t\to\infty} y_{1j}^{(N)}(t)$ exist and are equal, $j \neq 1$,*

(II) *for every $N = 1, 2, \ldots$, and all $t \geqslant U(N)$, $X_j^{(N)}(t)$ and $y_{1j}^{(N)}(t)$ are contained in an interval $[m_j^{(N)}, M_j^{(N)}]$, where*

$$\lim_{N\to\infty} m_j^{(N)} = \lim_{N\to\infty} M_j^{(N)} = \theta_j\,, \qquad j \neq 1.$$

In particular

$$\lim_{N\to\infty} \lim_{t\to\infty} X_j^{(N)}(t) = \lim_{N\to\infty} \lim_{t\to\infty} y_{1j}^{(N)}(t) = \theta_j\,, \qquad j \neq 1.$$

(III) *for every $N = 1, 2, \ldots$ and $j \neq 1$, the functions $f_j^{(N)} = y_{1j}^{(N)} - X_j^{(N)}$, $g_j^{(N)} = X_j^{(N)} - \theta_j$, and $\dot{y}_{1j}^{(N)}$ change sign at most once and not at all if $f_j^{(N)}(0)\, g_j^{(N)}(0) \geqslant 0$. Moreover, $f_j^{(N)}(0)\, g_j^{(N)}(0) > 0$ implies $f_j^{(N)}(t)\, g_j^{(N)}(t) > 0$ and $y_{1j}^{(N)}(t)$ is monotonic for all $t \geqslant 0$. Before proving the theorem, we illustrate* *its claim pictorially in Fig. 8 for the special case of a sequence of $G^{(N)}$'s with inputs*

$$I^{(N)}(t) = \sum_{k=0}^{N-1} J_1(t - k(w + W))$$

and

$$I_j^{(N)}(t) = \delta_{j2} \sum_{k=0}^{N-1} J_2(t - w - k(w + W)),$$

$j \neq 1$. *Such a sequence is obviously derived from the $G^{(\infty)}$ outstar in Corollary 3. The Figure compares two outstars $G^{(M)}$ and $G^{(N)}$ with $M \ll N$.*

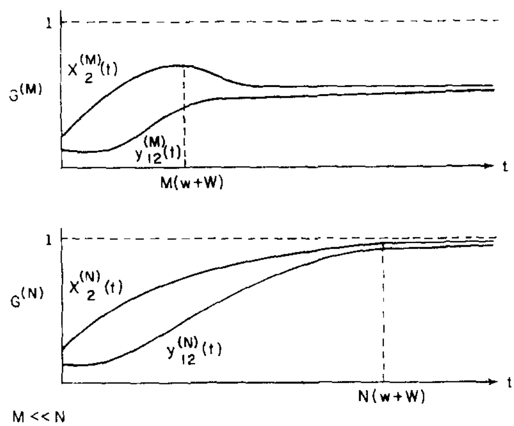

FIG. 8

PROOF. We carry out the proof in the case $\theta_j = \delta_{j2}$. The same method goes through in general, but it is more tedious. The idea of the proof is to try to divide the time interval $[0, \infty)$ of each $G^{(N)}$ into two parts $[0, \gamma_N)$ and $[\gamma_N, \infty)$, where $\lim_{N\to\infty} \gamma_N = \infty$. In $[0, \gamma_N)$, the functions of $G^{(N)}$ agree with those of $G^{(\infty)}$ and we can apply Theorem 4 to them. In $[\gamma_N, \infty)$, $G^{(N)}$ has an input-free border and we can apply Theorem 2 to its functions within this interval. This goal can be accomplished with but one technical reservation which appears in Case 2 below.

In every $G^{(N)}$, no input pulses reach the border during the time interval $[U(N), \infty)$, and we can therefore apply the results of Theorem 2 in this time interval. Thus, $y_{1j}^{(N)}(t)$ and $X_j^{(N)}(t)$ are monotonic in opposite senses and $| y_{1j}^{(N)}(t) - X_j^{(N)}(t) |$ decreases monotonically to zero for $t \geq U(N)$.

Letting

$$M_j^{(N)} = \max\{y_{ij}^{(N)}(U(N)), X_j^{(N)}(U(N))\}$$

and

$$m_j^{(N)} = \min\{y_{1j}^{(N)}(U(N)), X_j^{(N)}(U(N))\},$$

we conclude in particular that $y_{1j}^{(N)}(t)$ and $X_j^{(N)}(t)$ are contained in the interval $[m_j^{(N)}, M_j^{(N)}]$ for all $t \geqslant U(N)$. We must now distinguish two cases.

CASE 1. $U(N) \leqslant U_1(N)$. Clearly

$$X_j^{(N)}(t) = X_j^{(\infty)}(t) \qquad \text{and} \qquad y_{1j}^{(N)}(t) = y_{1j}^{(\infty)}(t)$$

for $t \in [0, U(N)]$. By Theorem 4,

$$\lim_{t\to\infty} X_j^{(\infty)}(t) = \lim_{t\to\infty} y_{1j}^{(\infty)}(t) = \delta_{j2},$$

$j \neq 1$. Since $U(N)$ increases to infinity with N,

$$\lim_{N\to\infty} X_j^{(N)}(U(N)) = \lim_{N\to\infty} y_{1j}^{(N)}(U(N)) = \delta_{j2}, \tag{30}$$

$j \neq 1$, and thus

$$\lim_{N\to\infty} M_j^{(N)} = \lim_{N\to\infty} m_j^{(N)} = \delta_{j2},$$

$j \neq 1$. In particular,

$$\lim_{N\to\infty} \lim_{t\to\infty} X_j^{(N)}(t) = \lim_{N\to\infty} \lim_{t\to\infty} y_{1j}^{(N)}(t) = \delta_{j2},$$

$j \neq 1$.

CASE 2. $U(N) > U_1(N)$. In this case, the conclusion of Case 1 still holds, but we need more information to reach it. By (16) and (17),

$$\dot{X}_2^{(N)} = A_1^{(N)}(y_{12}^{(N)} - X_2^{(N)}) + B_1^{(N)}(1 - X_2^{(N)}) \tag{16}$$

and

$$\dot{y}_{12}^{(N)} = C_1^{(N)}(X_2^{(N)} - y_{12}^{(N)}), \tag{17}$$

where $A_1^{(N)}$, $B_1^{(N)}$, and $C_1^{(N)}$ are nonnegative. If $\dot{y}_{12}^{(N)} \leqslant 0$, then by (17), $y_{12}^{(N)} - X_2^{(N)} \geqslant 0$ and so, by (16), $\dot{X}_2^{(N)} \geqslant 0$. If $\dot{y}_{12}^{(N)} \geqslant 0$, then by (17), $X_2^{(N)} \geqslant y_{12}^{(N)}$. In either case it is clear that $X_2^{(N)}(t)$ and $y_{12}^{(N)}(t)$ exceed $\min\{X_2^{(N)}(t_0), y_{12}^{(N)}(t_0)\}$ for any t and t_0 in $[U_1(N), U(N)]$ with $t \geqslant t_0$. In particular $X_2^{(N)}(U(N))$ and $y_{12}^{(N)}(U(N))$ exceed $\min\{X_2^{(N)}(U_1(N)), y_{12}^{(N)}(U_1(N))\}$. To show that (30) holds in this case, it therefore suffices to show that

$$\lim_{N\to\infty} X_2^{(N)}(U_1(N)) = \lim_{N\to\infty} y_{12}^{(N)}(U_1(N)) = 1.$$

This follows readily from Theorem 4 and the identities

$$X_2^{(N)}(U_1(N)) = X_2^{(\infty)}(U_1(N)) \qquad \text{and} \qquad y_{12}^{(N)}(U_1(N)) = y_{12}^{(\infty)}(U_1(N)).$$

In both cases, we have therefore shown that by taking N and then t sufficiently large, $X_2^{(N)}(t)$ and $y_{12}^{(N)}(t)$ can be brought as close to 1 as we please, and will thereafter *remain there*, even though no input pulses whatsoever occur at large times.

The conclusions of (III) follow simply by pasting together the results from Theorems 2 and 4. That is, we consider $G^{(N)}$ to be a $G^{(\infty)}$ for small times and an outstar with input-free border for large times.

This completes the proof that as N is taken increasingly large, the probability distributions $X_j^{(N)}$ and $y_{ij}^{(N)}$ of $G^{(N)}$ approximate the δ_{j2} distribution (or more generally any fixed probability distribution θ_j) with increasingly good accuracy for all large t.

(7B) The Outputs of Each $G^{(N)}$ Are Not a Good Index of the Stability of Its Probabilities

We will now show that in each $G^{(N)}$, the behavior of the output functions $x_j^{(N)}(t)$ can differ radically from the behavior of the ratios $X_j^{(N)}(t)$ as $t \to \infty$. In Theorem 5 we showed that the ratios $X_j^{(N)}(t)$ remain essentially fixed for large times t in each $G^{(N)}$. Now we show that the output functions $x_j^{(N)}(t)$ can be made to decay to zero at an exponential rate as $t \to \infty$. This contrast between ratios $X_j^{(N)}$ and outputs $x_j^{(N)}$, though technically simple to establish, will have a significant prediction theoretic meaning.

PROPOSITION 1. *In each $G^{(N)}$, the outputs $x_j^{(N)}$ decay exponentially to zero if $\alpha > 0$ and $u > 0$.*

PROOF. Since $G^{(N)}$ is input-free in $[\lambda(N), \infty)$, where

$$\lambda(N) = \max\{U_1(N), U(N)\},$$

(10) implies that

$$\dot{x}_1^{(N)}(t) = -\alpha x_1^{(N)}(t) \qquad \text{for} \qquad t \geqslant \lambda(N),$$

which proves the Proposition for $x_1^{(N)}$. Similarly, for $j \neq 1$ and $t \geqslant \lambda(N) + \tau$,

$$\dot{x}_j^{(N)}(t) = -\alpha x_j^{(N)}(t) + \beta x_1^{(N)}(t-\tau)\, y_{1j}^{(N)}(t),$$

and thus

$$-\alpha x_j^{(N)}(t) \leqslant \dot{x}_j^{(N)}(t) \leqslant -\alpha x_j^{(N)}(t) + \beta x_1^{(N)}(t-\tau).$$

Since the behavior as $t \to \infty$ of $x_1^{(N)}$ is already known, the Proposition follows immediately for all $x_j^{(N)}$, $j \neq 1$.

The assumptions $\alpha > 0$ and $u > 0$ are realized in those outstars whose outputs $x_j^{(N)}(t)$ eventually die out whenever the inputs $I_j^{(N)}(t)$ die out as well, and whose cross-correlations $z_{1j}^{(N)}(t)$ weight past $x_1(w-\tau)\, x_j(w)$ values with an exponentially decaying term $e^{-u(t-w)}$. This is the case for which our equations have a prediction theoretic interpretation.

(7C) The Effect of Fixed Ratios on Outputs

In (B) we showed that even though the ratios $X_j^{(N)}(t)$ remain essentially fixed for large t, the outputs $x_j^{(N)}(t)$ can decay exponentially to zero as $t \to \infty$. It therefore seems that the absolute magnitudes of the ratios and the outputs are completely unrelated as $t \to \infty$. This is not always true, but we must modify our outstars $G^{(N)}$ slightly to see this. For simplicity we again restrict attention to outstars $G^{(N)}$ with inputs

$$I_1^{(N)}(t) = \sum_{k=0}^{N-1} J_1(t - k(w+W))$$

and

$$I_j^{(N)}(t) = \delta_{j2} \sum_{k=0}^{N-1} J_2(t - w - k(w+W)).$$

To modify $G^{(N)}$, let $f_N(t)$ be any nonnegative and continuous function which is positive *only* in the interval $[U(N), \infty)$. Such an f_N is called ***admissible.*** Given any sequence $f = (f_1, f_2, \ldots, f_n, \ldots)$ of admissible f_N's, we shall now construct a sequence $G^{(1,f)}, G^{(2,f)}, \ldots, G^{(N,f)}, \ldots$ of outstars that is closely related to the sequence $G^{(1)}, G^{(2)}, \ldots, G^{(N)}, \ldots$ of outstars. For each

$N = 1, 2, \ldots, G^{(N,f)}$ is defined in terms of $G^{(N)}$ by the following prescriptions: (1) The initial data of $G^{(N,f)}$ is the same as that of $G^{(N)}$ (and hence that of $G^{(\infty)}$). (2) The input functions of $G^{(N,f)}$ are

$$I_1^{(N,f)} = I_1^{(N)} + f_N$$

and

$$I_j^{(N,f)} = I_j^{(N)}, \qquad j \neq 1.$$

We say that the sequence $G^{(1,f)}, G^{(2,f)}, \ldots$ is *derived* from f and $G^{(\infty)}$. Any derived sequence of this kind obeys the following theorem:

THEOREM 5(f). *Let $G^{(1,f)}, G^{(2,f)}, \ldots, G^{(N,f)}, \ldots$ be any sequence of outstars derived from any $G^{(\infty)}$ and any admissible f. Then all the conclusions of Theorem 5 hold for this sequence with superscripts "(N, f)" replacing superscripts "(N)".*

PROOF. Since $I_1^{(N,f)} = I_1^{(N)} + f_N$, the functions of $G^{(N,f)}$ agree with those of $G^{(N)}$ in $[0, U_1(N)]$. Since $I^{(N,f)} = I_j^{(N)}$, $j \neq 1$, $G^{(N,f)}$ has an input-free border in $[U(N), \infty)$. The rest of the proof is now just as in Theorem 5.

We now consider special choices of f which show some of the effects which the fixed ratios $X_j^{(N,f)}(t)$ can have on the outputs $x_j^{(N,f)}(t)$ for large t.

(1) $f_N(t) = J(t - \Lambda(N))$ *where* $\Lambda(N) \gg \lambda(N)$. For this choice of f_N, $G^{(N,f)}$ differs from $G^{(N)}$ only in the occurrence of an input pulse $J(t - \Lambda(N))$ at the source of $G^{(N,f)}$ at time $t = \Lambda(N)$. In particular, $G^{(N,f)}$ is the same as $G^{(N)}$ in $[0, \Lambda(N)]$, and so $G^{(N,f)}$ is input-free in $[\lambda(N), \Lambda(N)]$. By Proposition 1, the outputs $x_j^{(N,f)}(t)$ decay exponentially towards zero for $t \in [\lambda(N), \Lambda(N)]$. Since $\Lambda(N) \gg \lambda(N)$, we can assume that all the outputs $x_j^{(N,f)}$ are very small at time $t = \Lambda(N)$, and we write $x_j^{(N,f)}(\Lambda(N)) \cong 0$, $j = 1, 2, \ldots, n$. This is true for every $N \geq 1$.

By Theorem 5(f), we can by taking N sufficiently large guarantee that $y_{1j}^{(N,f)}(t)$ approximates δ_{j2} as closely as we wish for $t \geq \lambda(N)$. In particular, we can write $y_{1j}^{(N,f)}(t) \cong \delta_{j2}$ for $t \geq \Lambda(N)$. We are now ready to discuss the effects which the input pulse $f_N(t) = J(t - \Lambda(N))$ has on the outputs of $G^{(N,f)}$.

The first fact of interest is that $f_N(t)$ has essentially no effect whatsoever on the outputs $x_j^{(N,f)}$, $j \neq 1, 2$. By (11), we have for $t \geq \Lambda(N)$ and $j \neq 1, 2$ that

$$\begin{aligned} \dot{x}_j^{(N,f)}(t) &= -\alpha x_j^{(N,f)} + \beta x_1^{(N,f)}(t - \tau)\, y_{1j}^{(N,f)}(t) \\ &\cong -\alpha x_j^{(N,f)} + \beta x_1^{(N,f)}(t - \tau) \cdot 0 \\ &= -\alpha x_j^{(N,f)}. \end{aligned}$$

Thus

$$x_j^{(N,f)}(t) \cong e^{-\alpha(t-\Lambda(N))} x_j^{(N,f)}(\Lambda(N))$$
$$\cong 0.$$

By contrast, $f_N(t)$ has a substantial effect on the output $x_2^{(N,f)}$. By (11), we have for $t \geqslant \Lambda(N)$

$$\dot{x}_2^{(N,f)}(t) = -\alpha x_2^{(N,f)}(t) + \beta x_1^{(N,f)}(t-\tau)\, y_{12}^{(N,f)}(t)$$
$$\cong -\alpha x_2^{(N,f)}(t) + \beta x_1^{(N,f)}(t-\tau) \cdot 1,$$

and thus

$$x_2^{(N,f)}(t) \cong x_2^{(N,f)}(\Lambda(N))\, e^{-\alpha(t-\Lambda(N))} + \beta \int_{\Lambda(N)}^{t} e^{-\alpha(t-v)} x_1^{(N,f)}(v-\tau)\, dv$$
$$\cong \beta \int_{\Lambda(N)}^{t} e^{-\alpha(t-v)} x_1^{(N,f)}(v-\tau)\, dv. \tag{31}$$

But by (10) we have for $t \geqslant \Lambda(N)$ that

$$\dot{x}_1^{(N,f)}(t) = -\alpha x_1^{(N,f)}(t) + J(t-\Lambda(N)),$$

and thus

$$x_1^{(N,f)}(t) = x_1^{(N,f)}(\Lambda(N))\, e^{-\alpha(t-\Lambda(N))} + \int_{\Lambda(N)}^{t} e^{-\alpha(t-v)} J(v-\Lambda(N))\, dv$$
$$\cong \int_{\Lambda(N)}^{t} e^{-\alpha(t-v)} J(v-\Lambda(N))\, dv$$
$$= e^{-\alpha(t-\Lambda(N))} \int_0^{t-\Lambda(N)} e^{\alpha v} J(v)\, dv. \tag{32}$$

Substituting (32) into (31) gives for $t \in [\Lambda(N), \Lambda(N)+\tau]$ that

$$x_1^{(N,f)}(t) \cong 0$$

and for $t \geqslant \Lambda(N) + \tau$ that

$$x_2^{(N,f)}(t) \cong \beta \int_{\Lambda(N)+\tau}^{t} e^{-\alpha(t-v)} x_1^{(N,f)}(v-\tau)\, dv$$
$$\cong \beta e^{-\alpha(t-\Lambda(N)-t)} \int_0^{t-\Lambda(N)-\tau} dv \int_0^{v} e^{\alpha w} J(w)\, dw.$$

Thus $x_2^{(N,f)}(t)$ grows substantially in the interval $(\Lambda(N)+\tau, \Lambda(N)+\tau+\lambda)$, where $\lambda = \sup\{t : J(t) > 0\}$, and then decays once again at an exponential rate to zero.

We summarize these statements in Fig. 9.

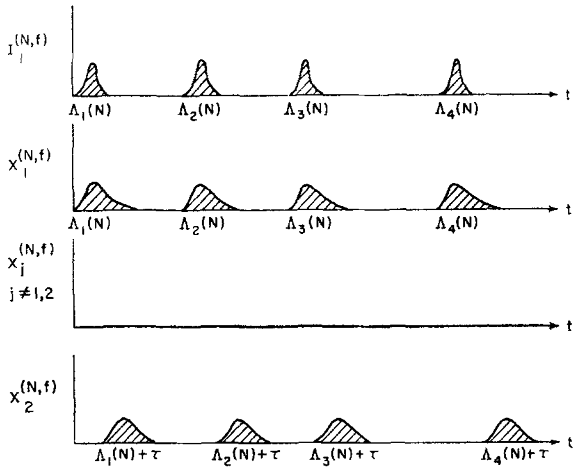

FIG. 9

These facts can be stated heuristically as follows. If only the vertices v_1 and v_2 are each perturbed periodically by N input pulses, where N is a large number, then a later test input pulse to v_1 creates a large output only from v_2. The inputs to the vertices v_1 and v_2 channel most of the mass $\sum_{k=2}^{n} y_{1k}^{(N,f)}(t)$ of the edges e_{1k} into the edge e_{12}, and then e_{12} channels a later test input pulse to v_1 along e_{12} and thence to v_2.

Suppose that the probability distribution δ_{j2} is replaced by an arbitrary probability distribution θ_j in the inputs $I_j^{(N,f)}$. Then the outputs from each v_j with $\theta_j > 0$ are affected by the input pulse f_N no matter how large N is taken. Clearly, for N taken sufficiently large, the output from v_i is approximately θ_i/θ_j times as large as the output from v_j.

The ratios $y_{12}^{(N,f)}(t)$ have a curious effect on the outputs $x_2^{(N,f)}(t)$ when both N and then t are taken sufficiently large. This effect depends on the following simple fact.

COROLLARY 4. *Let $G^{(N,f)}$ be chosen with $\theta_j = \delta_{j2}$. Then for all sufficiently large N and $t \geqslant m_2^{(N)}$, $y_{12}^{(N,f)}(t)$ is monotone increasing and $y_{1j}^{(N,f)}(t)$ is monotone decreasing, $j \neq 1$.*

PROOF. It suffices to consider $y_{12}^{(N,f)}(t)$. By Theorem 4, $y_{12}^{(\infty)}(t)$ is monotonic for all sufficiently large t. Since $y_{12}^{(\infty)}(t) \leqslant 1$ and $\lim_{t\to\infty} y_{12}^{(\infty)}(t) = 1$, $y_{12}^{(\infty)}(t)$ is monotone increasing for these t. Since also the sign of $\dot{y}_{12}^{(\infty)}(t)$ agrees with that of $X_2^{(\infty)}(t) - y_{12}^{(\infty)}(t)$, $X_2^{(\infty)}(t) > y_{12}^{(\infty)}(t)$ for sufficiently large t. Obviously $y_{12}^{(N,f)}(t) = y_{12}^{(\infty)}(t)$ and $X_2^{(N,f)}(t) = X_2^{(\infty)}(t)$ for $t \in [0, m_2^{(N)}]$ where

$\lim_{N\to\infty} m_2^{(N)} = \infty$. Thus $X_2^{(N,f)}(m_2^{(N)}) > y_2^{(N,f)}(m_2^{(N)})$ for all sufficiently large N. By (16) and (17) it now follows readily that $\dot{y}_{12}^{(N,f)}(t) > 0$ for $t \geqslant m_2^{(N)}$ and all sufficiently large N.

Using Corollary 4, we consider two derived sequences $G^{(N,f_1)}$ and $G^{(N,f_2)}$, $N = 1, 2,\ldots,$ which differ only in their choice of $\Lambda(N)$, say $\Lambda_1(N) \ll \Lambda_2(N)$. In both sequences we can assume that $x_j^{(N,f_i)}(t) \cong 0$ for $t \geqslant \Lambda_1(N)$; $i = 1, 2$; and $j = 1, 2,\ldots, n$. By Corollary 4, for all sufficiently large N and $t \geqslant \Lambda_1(N)$, $y_{12}^{(N,f_2)}(t) > y_{12}^{(N,f_1)}(t)$. By (10) and (11) this means that the output created by $J_1(t - \Lambda_2(N))$ at v_2 in $G^{(N,f_2)}$ will exceed the output created by $J_1(t - \Lambda_1(N))$ at v_2 in $G^{(N,f_1)}$. This fact is remarkable because $J_1(t - \Lambda_2(N))$ occurs later in time than $J_1(t - \Lambda_1(N))$, and thus after the outstar has had a greater opportunity to recover from the effect of prior inputs. Speaking heuristically, we therefore say that after sufficiently many input pulses have occurred at v_1 and v_2, a new input pulse to v_1 creates outputs from B which "spontaneously" seek the δ_{j2} distribution that the prior inputs have sought to establish. This "spontaneous facilitation" will, of course, be most evident when the gap $| y_{12}^{(N,f)}(U(N)) - X_2^{(N,f)}(U(N)) |$ is substantial.

(2) $f_N(t) = \sum_{k=1}^{\infty} J(t - \Lambda_k(N))$, $\lambda(N) \ll \Lambda_1(N) \ll \Lambda_2(N) \ll \cdots$. In this case, infinitely many input pulses occur at the source at large time separations $t = \Lambda_1(N), \Lambda_2(N),\ldots$. For this choice of f_N, we again readily conclude that the output functions $x_j^{(N,f)}$, $j \neq 1, 2$, are not affected by f_N if N is taken sufficiently large. Again the interest centers in $x_2^{(N,f)}$ for large N.

We can treat $x_2^{(N,f)}$ just as we did in case (1) for times $t \in [0, \Lambda_2(N)]$. Treating $\Lambda_1(N)$ as the $\Lambda(N)$ of case (1), we conclude that $x_2^{(N,n)}$ grows substantially in the interval $(\Lambda_1(N) + \tau, \Lambda_1(N) + \tau + \lambda)$ and then decays

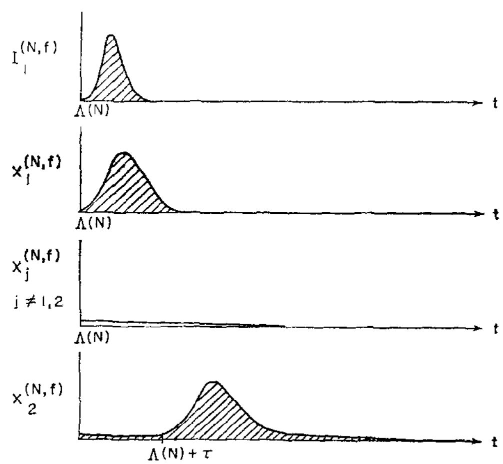

FIG. 10

exponentially towards zero in $[A_1(N) + \tau + \lambda, A_2(N)]$. Since $A_2(N) \gg A_1(N)$, we can suppose $x_2^{(N,f)}(A_2(N)) \simeq 0$. Now we iterate this process. We treat $A_2(N)$ as the $A(N)$ of case (1) and $A_1(N)$ as the $\lambda(N)$ of case (1). We conclude that $x_2^{(N,f)}$ grows substantially in the interval $(A_2(N) + \tau, A_2(N) + \tau + \lambda)$ and then decays exponentially towards zero in $[A_2(N) + \tau + \lambda, A_3(N)]$. This process is iterated infinitely often, and we arrive at Fig. 10. The heuristic point of this example is that we can perturb the source as often as we wish with input pulses without distorting the output, just so long as the rate with which the pulses occur is sufficiently slow. This remark must of course be qualified by the spontaneous improvement effect noted in (1).

(7D) The Nonlinear Trend in the Individual Outputs is Not Seen in the Linear Average Output

In Sections (7A)-(7C), we have shown that there exists a distinctive trend in the outputs of a sequence $G^{(1,f)}, G^{(2,f)},..., G^{(N,f)},...$ of outstars if, for example, we let $f_N(t) = J(t - A(N))$, where $A(N) \gg \lambda(N)$. This trend is particularly evident if we let $G^{(\infty)}$ have initial data of the form $z_{1j}(0) = \delta > 0$, and $x_j(0) = \gamma$, for all $j \neq 1$, and choose $\theta_j = \delta_{j2}$. Then in every $G^{(N,f)}$, the output from each vertex of the border is the same at time $t = 0$, and we say that the output is *uniformly distributed* at time $t = 0$. In $G^{(1,f)}$, the distribution of outputs from the border never deviates too far from this uniform distribution since only one input pulse reaches v_1 and v_2. In $G^{(2,f)}$, the distribution of outputs from the border is slightly more peaked at v_2 for times $t \geq A(2)$ than it is in $G^{(1,f)}$ for times $t \geq A(1)$. By the time we reach a $G^{(N,f)}$ for which N is very large, practically all the output from the border comes from vertex v_2 for times $t \geq A(N)$. We diagram this trend in an idealized way in Fig. 11, where we have set $w = \tau$ for simplicity.

We now ask how much of this striking trend is visible in the average output

$$x^{(N,f)} = \frac{1}{n} \sum_{k=1}^{n} x_k^{(N,f)}$$

of each outstar $G^{(N,f)}$, $N = 1, 2,...$. We will show that this trend need not appear at all in these averages for large times.

PROPOSITION 2. *For any*

$$f_N(t) = \sum_{k=1}^{R} J(t - A_k(N)),$$

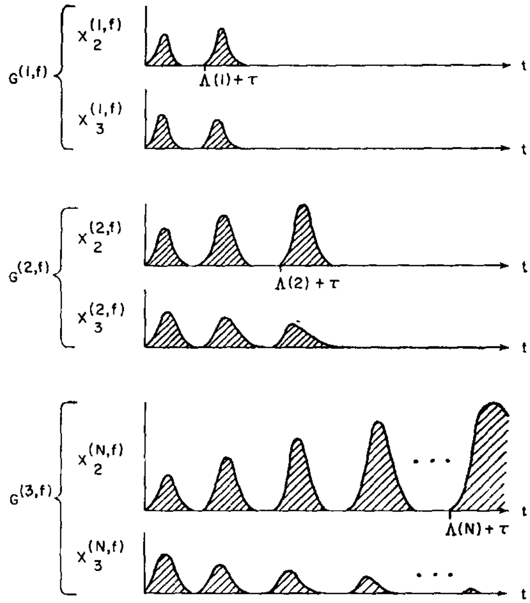

FIG. 11

where

$$\Lambda_{k+1}(N) - \Lambda_k(N) = \Lambda_{k+1}(M) - \Lambda_k(M) \gg 0,$$

and R is any nonnegative integer, including ∞,

$$x^{(M,f)}(t + \Lambda_1(M)) \cong x^{(N,f)}(t + \Lambda_1(N))$$

for all $t \geqslant 0$ and all M, $N = 1, 2, \ldots$.

PROOF. We prove the proposition only for the case $R = 1$. The generalization to other values of R will then be obvious. The proof relies on the fact from Corollary 1 that the average $x^{(N,f)}$ obeys a linear equation which is independent of the probability distribution $y_{1j}^{(N,f)}$. We omit the subscript "1" in $\Lambda_1(N)$ for simplicity. From Corollary 1 we find that

$$\left(\sum_{j\neq 1} x_j^{(N,f)}\right)^{\boldsymbol{\cdot}} = -\alpha\left(\sum_{j\neq 1} x_j^{(N,f)}\right) + \beta x_1^{(N,f)}(t-\tau).$$

Since also

$$\dot{x}_1^{(N,f)} = -\alpha x_1^{(N,f)} + J(t - \Lambda(N)),$$

$$\dot{x}^{(N,f)} = -\alpha x^{(N,f)} + \frac{\beta}{n} x_1^{(N,f)}(t-\tau) + \frac{1}{n} J(t - \Lambda(N)). \tag{33}$$

Integrating (33) gives for $t \geqslant \Lambda(N)$,

$$x^{(N,f)}(t) = x^{(N,f)}(\Lambda(N))\, e^{-\alpha(t-\Lambda(N))} + \frac{1}{n}\int_{\Lambda(N)}^{t} e^{-\alpha(t-v)}[\beta x_1^{(N,f)}(v-\tau) + J(v-\Lambda(N))]\, dv$$

$$\cong \frac{\beta}{n}\int_{\Lambda(N)}^{t} e^{-\alpha(t-v)} x_1^{(N,f)}(v-\tau)\, dv + \frac{1}{n} e^{-\alpha(t-\Lambda(N))}\int_0^{t-\Lambda(N)} e^{\alpha v} J(v)\, dv,$$

and since

$$x_1^{(N,f)}(t) \cong e^{-\alpha(t-\Lambda(N))}\int_0^{t-\Lambda(N)} e^{\alpha v} J(v)\, dv,$$

$$x^{(N,f)}(t) \cong \frac{\beta}{n}\chi(t-\Lambda(N)-\tau)\int_0^{t-\Lambda(N)-\tau} dv \int_0^{v} e^{\alpha w} J(w)\, dw + \frac{1}{n} e^{-\alpha(t-\Lambda(N))}\int_0^{t-\Lambda(N)} e^{\alpha v} J(v)\, dv.$$

Thus for any $t \geqslant 0$,

$$x^{(N,f)}(t+\Lambda(N)) \cong x^{(M,f)}(t+\Lambda(M)).$$

The heuristic point of Proposition 2 can be stated for the case in which $R = \infty$ in the following way. Suppose that an experimentalist wants to find out how outstars work by collecting data from them. A standard rule of prudence when confronted with an unknown system is to first study the long time average output of the system. Given the outstar $G^{(N,f)}$, this average is

$$\frac{1}{t}\int_0^t x^{(N,f)}(v)\, dv,$$

where $t \gg 0$. The experimentalist will readily observe that each average output $x^{(N,f)}$ obeys a simple linear equation. He will also note that

$$\frac{1}{t}\int_0^t x^{(N,f)}(v)\, dv \cong \frac{1}{t}\int_0^t x^{(M,f)}(v)\, dv$$

for t sufficiently large and all $M, N = 1, 2, \ldots$, since

$$x^{(N,f)}(t+\Lambda(N)) \cong x^{(M,f)}(t+\Lambda(M)),$$

and an infinite amount of mass f_N reaches the source vertex for times $t \geqslant \Lambda(N)$ in every $G^{(N,f)}$. A plausible inference from this data is that all the outstars

$G^{(1,f)},\ldots, G^{(N,f)},\ldots$ obey a linear equation and that these outstars are essentially copies of one another. Both of these conclusions are totally wrong!

8. Eliminating Background Noise in Outputs: The Entropy of an Outstar

In this section, we will show how the outputs of an outstar can be modified to eliminate background noise. We again consider the special case of N-truncations $G^{(N)}$ with inputs

$$I_1^{(N)}(t) = \sum_{k=0}^{N-1} J_1(t - k(w + W))$$

and

$$I_j^{(N)}(t) = \delta_{j2} \sum_{k=0}^{N-1} J_2(t - w - k(w + W)), \qquad j \neq 1.$$

For simplicity, we also require that each $G^{(N)}$ have an *initially uniform border*; i.e., $z_{1j}(0) = \delta > 0$ and $x_j(0) = \gamma > 0$, $j \neq 1$.

The need to modify the outputs $x_j^{(N)}(t)$ is suggested by Theorem 5. This theorem says that

$$\lim_{N\to\infty} \lim_{t\to\infty} X_j^{(N)}(t) = \lim_{N\to\infty} \lim_{t\to\infty} y_{1j}^{(N)}(t) = \delta_{j2}$$

or that the outputs from $G^{(N)}$ come increasingly from v_2 as N and then t are taken large. By contrast, if $x_j^{(N)}(t_0) > 0$ for any $t_0 \geqslant 0$, then $x_j^{(N)}(t) > 0$ for every $t \geqslant t_0$. Thus outputs from vertices v_3, v_4,..., v_n cannot be entirely eliminated in finite time and produce background noise. We now introduce one way of modifying the outputs to eliminate this noise and several other difficulties.

Given any probability distribution $p = (p_1, p_2, \ldots, p_{n-1})$, let the *entropy* $H(p)$ of p be defined by

$$H(p) = - \sum_{k=1}^{n-1} p_k ln_2 p_k ,$$

where it is understood that $0 ln_2 0 = 0$. This concept of entropy is familiar from information theory, and it provides a rigorous measure of the amount of information in a scheme of events [8]. Using this familiar notion of entropy, we can define two kinds of entropy in any outstar $G^{(N)}$, $N = 1, 2, \ldots, \infty$. Let

$$H_X^{(N)}(t) = H(X_2^{(N)}(t), \ldots, X_n^{(N)}(t))$$

be the *vertex* (or *state*) *entropy* of the border of $G^{(N)}$ at time t. Let

$$H_y^{(N)}(t) = H(y_{12}^{(N)}(t),..., y_{1n}^{(N)}(t))$$

be the *edge* (or *interaction*) *entropy* of the border of $G^{(N)}$ at time t. These entropy functions have the following properties.

PROPOSITION 3. *Let $G^{(\infty)}$ be any outstar with an initially uniform border and input functions*

$$I_1^{(\infty)} = \sum_{k=0}^{\infty} J_1(t - k(w + W))$$

and

$$I_j^{(\infty)} = \delta_{j2} \sum_{k=0}^{\infty} J_2(t - w - k(w + W)), \qquad j \neq 1.$$

Then the state entropy $H_X^{(\infty)}$ and the interaction entropy $H_y^{(\infty)}$ of $G^{(\infty)}$ attain the maximum entropy $ln_2(n-1)$ at time $t = 0$ and approach the minimum entropy 0 as $t \to \infty$. Moreover, $H_y^{(\infty)}(t)$ decreases monotonically from maximal entropy to minimal entropy as $t \to \infty$, and

$$\lim_{N\to\infty} \lim_{t\to\infty} H_X^{(N)}(t) = \lim_{N\to\infty} \lim_{t\to\infty} H_y^{(N)}(t) = 0.$$

PROOF. The maximum entropy of $H(p)$ is $ln_2(n-1)$ and is attained when $p = [1/(n-1),..., 1/(n-1)]$. By hypothesis, $X_j^{(\infty)}(0) = y_{1j}^{(\infty)}(0) = 1/(n-1)$, $j \neq 1$. Thus

$$H_X^{(\infty)}(0) = H_y^{(\infty)}(0) = ln_2(n-1).$$

The entropy $H(p)$ is also a continuous function of p whose minimum value 0 is attained when $p = (1, 0, 0,..., 0)$ (say). By Corollary 3,

$$\lim_{t\to\infty} X_j^{(\infty)}(t) = \lim_{t\to\infty} y_{1j}^{(\infty)}(t) = \delta_{j2}.$$

Thus

$$\lim_{t\to\infty} H_X^{(\infty)}(t) = \lim_{t\to\infty} H_y^{(\infty)}(t) = 0.$$

By Theorem 5, we also know that

$$\lim_{N\to\infty} \lim_{t\to\infty} H_X^{(N)}(t) = \lim_{N\to\infty} \lim_{t\to\infty} H_y^{(N)}(t) = 0,$$

To show that $H_y^{(\infty)}(t)$ decreases monotonically, note that

$$\frac{d}{dp_1} H\left(p_1, \frac{1}{n-2}(1-p_1), \frac{1}{n-2}(1-p_1), \ldots, \frac{1}{n-2}(1-p_1)\right)$$

$$= \frac{\log_e\left(\frac{1-p_1}{(n-2)p_1}\right)}{\log_e(n-1)} \begin{cases} \leqslant 0 & \text{if} \quad p_1 \geqslant \frac{1}{n-1} \\ > 0 & \text{if} \quad p_1 < \frac{1}{n-1}. \end{cases} \tag{34}$$

Since $G_y^{(\infty)}$ has a uniformly distributed border at $t = 0$ and $I_j \equiv 0, j \neq 1, 2,$

$$y_{13}^{(\infty)} \equiv y_{13}^{(\infty)} \equiv \cdots \equiv y_{1n}^{(\infty)} \qquad \text{and} \qquad y_{1j}^{(\infty)} = \frac{1}{n-2}(1 - y_{12}^{(\infty)}), \qquad j \neq 1, 2.$$

Thus

$$H_y^{(\infty)}(t) = H(y_{12}^{(\infty)}, \frac{1}{n-2}(1 - y_{12}^{(\infty)}), \ldots, \frac{1}{n-2}(1 - y_{12}^{(\infty)})).$$

Differentiating $H_y^{(\infty)}$ therefore gives

$$\dot{H}_y^{(\infty)}(t) = \frac{d}{dy_{12}^{(\infty)}} H\left(y_{12}^{(\infty)}, \frac{1}{n-2}(1 - y_{12}^{(\infty)}), \ldots, \frac{1}{n-2}(1 - y_{12}^{(\infty)})\right) \dot{y}_{12}^{(\infty)}. \tag{35}$$

To calculate the sign of $\dot{y}_{12}^{(\infty)}$ consider Theorem 4. Since $y_{12}^{(\infty)}(0) = X_2^{(\infty)}(0)$, $y_{12}^{(\infty)}(t)$ is a monotonic function for all $t \geqslant 0$. Since

$$y_{12}^{(\infty)}(0) = \frac{1}{n-1} \qquad \text{and} \qquad \lim_{t \to \infty} y_{12}^{(\infty)}(t) = 1,$$

$y_{12}^{(\infty)}(t)$ is monotone increasing and $y_{12}^{(\infty)}(t) \geqslant 1/(n-1)$. By (34) and (35), we therefore find that $\dot{H}_y^{(\infty)}(t) \leqslant 0$, or that $H_y^{(\infty)}(t)$ is monotone decreasing.

The decrease of $H_y^{(\infty)}(t)$ from maximum to minimum "lack of information" as $t \to \infty$ will be associated in Part II with the increase in $G^{(\infty)}$ from minimum to maximum "learning" as $t \to \infty$.

The *modified output* $0_j^{(N)}(t, \Gamma)$ of $x_j^{(N)}(t)$ is defined in terms of the entropy as follows.

$$0_j^{(N)}(t, \Gamma) = \max\{x_j^{(N)}(t)\, \bar{H}_X^{(N)}(t) - \Gamma, 0\},$$

where

$$\bar{H}_X^{(N)}(t) = 1 - \frac{H_X^{(N)}(t)}{\ell n_2(n-1)},$$

and Γ is a fixed positive number called the *output threshold*. We list several properties of $0_j^{(N)}(t, \Gamma)$ in a $G^{(N)}$ with initially uniform border and provide a brief intuitive interpretation of each property.

(1) $0_j^{(N)}(0, \Gamma) = 0$ for *all* $\Gamma > 0$, no matter how large δ and γ are. No vertex is "preferred" at $t = 0$; thus no output occurs.

(2) Only $0_2^{(N)}(t, \Gamma)$ ever becomes positive if Γ is chosen sufficiently large. That is, only the "favored" edge e_{12} ever generates an output.

(3) $0_2^{(\infty)}(t + w + W, \Gamma) \geqslant 0_2^{(\infty)}(t, \Gamma)$, $t \geqslant 0$. As inputs to v_1 and v_2 concentrate more of the mass $\sum_{k \neq 1} y_{1k}^{(\infty)}$ at e_{12} in successive input periods, so too does the output from v_2 become stronger.

(4) If $x_j^{(N)}(t) = 0$, then $0_2^{(N)}(t, \Gamma)$ is independent of $x_j^{(N)}(t)$. A vertex which has never been "created" by an input influences no output.

$$(5) \qquad \lim_{t \to \infty}[0_2^{(\infty)}(t, \Gamma) - \max(x_2^{(\infty)}(t) - \Gamma, 0)] = 0.$$

The output from v_2 becomes independent of the outputs $x_j^{(\infty)}$, $j \neq 1, 2$, as all edges e_{1j} lose mass to e_{12}. In particular, when $\Gamma = 0$, the modified output converges to the unmodified output as $t \to \infty$.

In summary, the modified output $0_2^{(\infty)}(t, \Gamma)$ behaves essentially like the unmodified output $x_2^{(\infty)}(t)$ when only $y_{12}^{(\infty)}(t)$ is large, but $0_j^{(\infty)}(t, \Gamma) \cong 0$ when all $y_{1j}^{(\infty)}(t)$ are comparable, even when the $x_j^{(\infty)}(t)$'s are very large.

PART II PREDICTION THEORETIC INTERPRETATION

1. Introduction

We now give the results of Part I a prediction theoretic interpretation. Our goal is to construct laws for a machine $\mathscr{M}$ which can be taught to predict the event B whenever the event A occurs. This goal can be stated in several related ways. We can say that we wish to teach the machine that the transition $A \to B$ is correct, or that we wish to teach the machine the list AB.Phrased in this way, our task can be described by analogy with the task of teaching lists of letters to an idealized human subject, who will henceforth be denoted by $\mathscr{S}$. Suppose that we wish to teach $\mathscr{S}$ the list of letters AB. A standard way of doing this is to repeat the list AB to $\mathscr{S}$ several times. To find out if $\mathscr{S}$ has learned the list as a result of these list repetitions, the letter A alone is then said to $\mathscr{S}$. If $\mathscr{S}$ responds by saying the letter B in return, and $\mathscr{S}$ does this whenever A alone is said, then we have good evidence that $\mathscr{S}$ has indeed learned the list AB. Thus $\mathscr{S}$ learns to predict the event B whenever the event A occurs as a result of repeated presentations of the list AB.

In this section, we suggest one way of translating this intuitive idea of learning into formal terms. We can easily think of several desirable properties which a machine that learns a list of events in this way might profitably have. We state these properties here in a somewhat colorful language to aid the reader in comparing and contrasting his intuitive concepts of learning with the particular formal translation table that we will set down for these properties. The translation table that we shall provide is a very special one, to be sure, since it is intended to deal with the particularly simple case of an outstar.

(1) Practice Makes Perfect. The more often the list AB of events is repeated to the machine $\mathscr{M}$, the better becomes $\mathscr{M}$'s prediction of B given A. Moreover, if the list AB is repeated indefinitely often, then $\mathscr{M}$'s prediction of B given A comes as close as we wish to a perfect prediction.

(2) An Isolated System Suffers No Memory Loss. If we succeed in teaching the list AB to $\mathscr{M}$ to a given degree of accuracy, then $\mathscr{M}$ remembers the list with approximately this accuracy just so long as no new teaching occurs.

(3) An Isolated System Remembers and Sometimes Facilitates Its Memory without Continually Practicing. In everyday life, it is a commonplace experience that facts can be remembered for a substantial time in the absence of continual overt practice. We shall construct a machine that also has a good memory even when it does not practice. Indeed, its memory sometimes spontaneously improves even without practice (i.e., "reminiscence" occurs, [9], p. 509).

(4) The Act of Making a Correct Prediction Can Reoccur Indefinitely Often without Retraining. Suppose that $\mathscr{M}$ knows the list AB of events. It would be most unpleasant if the very act of predicting B, given A, erased the record within $\mathscr{M}$ that B is indeed the correct reply to A. If this were true, we would have to reteach the list AB every time a correct prediction occurred. In the present system, the act of recall can occur as many times as we please without requiring the retraining of $\mathscr{M}$.

Properties (1)-(4) show that once a list AB of events is taught to the machine $\mathscr{M}$, retention of the list is quite stable. The next property shows that this stability does not prevent $\mathscr{M}$ from adapting to new experiences.

(5) All Errors Can Be Corrected. Suppose after $\mathscr{M}$ learns the list AB it is found that really the event C should follow the event A. Then B is, by fiat, an *error* whenever it follows A. We shall see that this error can always be corrected if $\mathscr{M}$ then practices the list AC sufficiently often.

We now make properties (1)-(5) rigorous by translating them into theorems of Part I about outstars.

2. The Machine

The machine $\mathscr{M}$ we construct here obeys the equations (10)-(13) of an outstar. Once this machine is understood, the same basic concepts can be applied to a system given by any semistochastic matrix P, as defined in Section 2. $\mathscr{M}$ consists of n states, namely the n vertices v_i of the outstar, and these states interact with one another along the directed edges e_{1i}. The machine $\mathscr{M}$ is manipulated by an experimenter $\mathscr{E}$ whose goal is to teach $\mathscr{M}$ to predict the event B given the event A. The experimental manipulations created by $\mathscr{E}$ are represented by the input vector function $C = (I_1, I_2, ..., I_n)$. The outputs which these manipulations produce are represented by the output vector function $X = (x_1, x_2, ..., x_n)$. In particular, the input function I_j represents the total history of experimental manipulations performed on state v_j.

Suppose for example that

$$I_1 = \sum_{k=0}^{\infty} J_1(t - k(w + W))$$

and

$$I_j = \theta_j \sum_{k=0}^{\infty} J_2(t - w - k(w + W)), \qquad j \neq 1.$$

Then $\mathscr{M}$ is an outstar of type $G^{(\infty)}$, as treated in Corollary 3. Each function $J_1(t - k(w + W))$ and $\theta_j J_2(t - w - k(w + W))$ signifies the occurrence of an experimentally created event. $J_1(t - k(w + W))$ is an event which begins at the source vertex v_1 at time $t = k(w + W)$ and lasts until time $t = k(w + W) + \lambda_1$. The function $I_1 = \sum_{n=0}^{\infty} J_1(t - k(w + W))$ signifies the occurrence at the source v_1 of a periodic succession of identical events with the waveform J_1 at the times $t = 0, w + W, 2(w + W),...$. These events each last λ_1 time units. Similarly, the function

$$I_j = \theta_j \sum_{n=0}^{\infty} J_2(t - w - k(w + W))$$

signifies the occurrence at the border vertices of a periodic succession of identical events with waveform $\theta_j J_2$ at the times $t = w, 2w + W, 3w + 2W,...$.

These events each last λ_2 time units. Every vertex v_j of the border receives a periodic succession of events whose waveform is identical except for the weights θ_j. That is, the experimenter distributes a fraction θ_j of the waveform J_2 to each vertex v_j of the border at periodic intervals. Corollary 3 assures the experimenter that his unceasing labors do not go unnoticed by the outstar. Indeed the normalized vertex functions X_j and the normalized edge functions y_{1j} are both responsive to the experimenters's choice of weights θ_j and gradually adopt these weights as their own no matter what their weights were initially.

3. Repeating the List AB N Times

Now that we know what an event means in an outstar, it is simple to translate into formal terms the intuitive idea of presenting a list AB of events to the outstar N times. Firstly we must assign a state of the outstar to each symbol of an event. If for example we are given twenty-six symbols A, B, C,..., Z, then we assign v_1 to A, v_2 to B, v_3 to C, and so on down to v_{26} and Z. Given this assignment of symbols to states, suppose that an experimenter wishes to teach an outstar to predict B given A. He must indicate to the outstar in some way that B is the "correct" successor of A. He does this by repeating the desired sequence AB several times. The only way to say a sequence AB to an outstar is to create perturbations at the vertices v_1 and v_2 which stand for A and B, respectively. Thus one occurrence of the sequence AB is translated into an outstar's mechanism by the arrival of an input pulse $J(t - k(w + W))$ at v_1 and of an input pulse $J(t - w - k(w + W))$ at v_2 w time units later. N periodic presentations of the sequence AB, starting at time zero, is translated into an outstar's mechanism as an input function

$$I_1^{(N)}(t) = \sum_{k=0}^{N-1} J(t - k(w + W))$$

for vertex v_1, an input function

$$I_2^{(N)}(t) = \sum_{k=0}^{N-1} J(t - w - k(w + W))$$

for vertex v_2, and input functions $I_j^{(N)} \equiv 0$ for all other vertices $j \neq 1, 2$; that is, as an N-trunction $G^{(N)}$. $G^{(N)}$ is thus the outstar which results when AB is repeated N times at a fixed rate by the experimenter.

To test whether or not $G^{(N)}$ has learned to predict B given A, the experimenter presents A to $G^{(N)}$ at a later time and sees whether or not $G^{(N)}$ knows that B is the correct prediction. That is, the experimenter creates an input pulse at v_1 and waits to see if the output created in this way comes only from v_2. As soon as A occurs, however, the outstar is no longer of type $G^{(N)}$.

Suppose, for example, that A occurs in $G^{(N)}$ at time $t = A(N)$. This means that the input pulse $f_N(t) = J(t - A(N))$ occurs at the source v_1. The total input to the source is therefore $\sum_{k=0}^{N-1} J(t - k(w + W)) + f_N(t)$. This is the input of an outstar of type $G^{(N,f)}$. This outstar of type $G^{(N,f)}$, where $f_N(t) = J(t - A(N))$, is thus a machine subjected to N presentations of the list AB followed by a single presentation of A on a test trial.

4. "Practice Makes Perfect"

Suppose now that a machine of type $G^{(N,f)}$ is given. That is, the experimenter has presented AB to the machine N times and then presents A alone. The experimenter wants the machine to predict B after A occurs. This means that the output from the border created by f_N ought to come only from v_2 if the machine knows the list AB. Theorem 5(f) shows that the output comes increasingly from v_2 as N increases. This means that the machine learns to predict B given A with ever greater precision as it receives ever more trials on which to practice the sequence AB. In this sense, the outputs at large times from the sequence $G^{(1,f)}, G^{(2,f)},..., G^{(N,f)},...$ of outstars, where $f_N(t) = J(t - A(N))$ exemplify the proverb "practice makes perfect" in our formal translation table.

This proverb is the first property stated in Section 1. The second property is that "an isolated system suffers no memory loss." An "isolated system" is manifestly one that is input-free. Property 2 can thus be stated formally as follows. The probability distributions $X_j^{(N,f)}$ and $y_{1j}^{(N,f)}$ of an outstar of type $G^{(N,f)}$ remain essentially fixed for all large t. This is proved in Theorem 5(f). The third property is that "an isolated system remembers without continually practicing." This is the statement of Section 5 that the probability distributions $X_j^{(N)}$ and $y_{1j}^{(N)}$ remain fixed even as the outputs $x_j^{(N)}$ decay exponentially to zero. A similar remark holds in an outstar of type $G^{(N,f)}$. Consider an experimenter who is studying $G^{(N,f)}$ for times $t \in [\lambda(N), A(N)]$; that is, after AB has occurred N times and before A alone occurs. He will certainly observe the rapid exponential decay of all the outputs $x_j^{(N,f)}$ and might well therefore be led to conclude that the effects of saying AB to the outstar N times wear off rapidly. The outstar provides no overt evidence (e.g., no "overt practice") to the experimenter during this time that any record whatever of his having presented AB endures. Nonetheless, by Theorem 5(f), shortly after A is presented to the outstar at time $t = A(N)$, the output is produced by B alone if AB has been said sufficiently often in the past. We also observed in Corollary 4 that if N is taken sufficiently large, $y_{12}^{(N,f)}(t)$ *increases* for $t \geqslant m_2^{(N)}$. On a later test trial, therefore, the outstar's memory can be better than it was immediately after training, even though no overt practice intervenes. This effect is most pronounced when N is large but not so large that $| y_{12}^{(N,f)}(U(N)) - X_2^{(N,f)}(U(N)) |$ is small; i.e., after "moderate" amounts of practice.

The fifth property of Section 1 is that "all errors can be corrected." This property is discussed in the next section.

5. Error Correction and Global Theorems

Suppose that an experimenter has taught an outstar the list AB by presenting this list N times to the outstar at a fixed rate. If N is taken sufficiently large, Theorem 5 guarantees that the list can be learned to an arbitrary degree of accuracy. After accomplishing this goal, suppose that another experimenter comes upon the outstar. This experimenter wishes to teach the oustar the list AC. He tests whether the outstar already knows this list by presenting the test pulse $\tilde{J}_1(t - A_1)$ to vertex v_1 at time $t = A_1$. The output created in this way comes almost exclusively from B. Because this experimenter *wants* C to be the output instead of B, he interprets the output from B as an *error*. To correct this error, he begins at time $t = A_2$ to present the list AC to the outstar M times at a fixed rate of speed, where M is chosen sufficiently large to offset the previous N occurrences of the "incorrect" list AB.

The input history of this outstar can be written as

$$I_1(t) = \sum_{k=0}^{N-1} J_1(t - k(w_1 + W_1)) + \tilde{J}_1(t - A_1) + \sum_{k=0}^{M-1} \tilde{J}_1(t - A_2 - k(w_2 + W_2)),$$

$$I_2(t) = \sum_{k=0}^{N-1} J_2(t - w_1 - k(w_1 + W_1)),$$

$$I_3(t) = \sum_{k=0}^{M-1} \tilde{J}_3(t - A_2 - w_2 - k(w_2 + W_2)),$$

$$I_j(t) \equiv 0, \qquad j \neq 1, 2, 3,$$

where J_i and $\tilde{J}_i$ are input pulses that are positive in $(0, \lambda_i)$ and $(0, \tilde{\lambda}_i)$, respectively, $i = 1, 2, 3$, and w_i and W_i are positive numbers, $i = 1, 2$. The basic question is: by repeating AC a sufficiently large number of times M, can the record of previous occurrences of AB N times be erased from within the outstar? The answer is "yes." This is because of two facts: (1) the outstar has positive data at time $t = A_1 + \tilde{\lambda}_1$ and in the interval $[A_1 + \tilde{\lambda}_1, \infty)$ the inputs are

$$I_1^{(t)} = \sum_{k=0}^{M-1} \tilde{J}_1(t - A_2 - k(w_2 + W_2)),$$

$$I_3(t) = \sum_{k=0}^{M-1} \tilde{J}_3(t - A_2 - w_2 - k(w_2 + W_2)),$$

and

$$I_j \equiv 0, \qquad j \neq 1, 3;$$

(2) Theorem 5 is true for any outstar with positive data and inputs of the form given in (1).

Thus the possibility of correcting all errors in an outstar depends on two facts: (1) *invariance of the set of all initial data under inputs:* no matter what inputs occur in a finite time interval, just so long as they are nonnegative the functions of the outstar will remain positive; and (2) *the limit of ratios of solutions is independent of the initial data:* Theorem 5(f) is true no matter what the initial data is, just so long as it is positive.

This discussion completes our formal translation table of the heuristic properties (1)-(5) of Section 1 in the case of an outstar. We remark in passing that properties (1)-(5) do not hold for systems characterized by arbitrary semistochastic matrices P. The way in which a system learns to predict depends in an essential way on the matrix P that characterizes it; i.e., on its "geometry." We will provide another example of this fact in the next paper in this series.

6. Improving the Accuracy of Predictions

In Section 5 of Part I, we showed how modifying the outputs of an outstar by using the notion of entropy removes background noise from these outputs. This modification of the output can now be interpreted prediction theoretically. We list the five properties of Section 8 in prediction theoretic terminology for the convenience of the reader.

(1) If no event B, C,... is a preferred prediction, given the event A, then no prediction is made.

(2) A prediction B given A is made only when the machine prefers the transition $A \to B$ on the basis of past experience.

(3) The strength of the prediction B given A increases as the machine's preference for the transition $A \to B$ increases with more practice of the list AB.

(4) If an event C has never occurred in the machine's past, it does not influence the prediction of B given A.

(5) As the list AB is practiced indefinitely often, even events C which have occurred in the machine's remote past gradually lose their influence on the prediction of B given A.

7. Information vs. Learning

In Proposition 3 we showed that the interaction entropy $H_y^{(\infty)}(t)$ decreases monotonically from maximal to minimal entropy as $t \to \infty$ in any $G^{(\infty)}$ with an initially uniform border. From an information theoretic viewpoint, this fact means that the "uncertainty" or "lack of information" in the scheme of events described by the probabilities $(y_{\infty}^{(12)}(t),..., y_{1n}^{(\infty)}(t))$ decreases as $t \to \infty$. A parallel interpretation along prediction theoretic lines is also available; namely, $G^{(\infty)}$ "learns" that v_2 follows v_1 with ever increasing accuracy as $t \to \infty$. At $t = 0$, $G^{(\infty)}$ is in a state of minimal "learning" since all $y_{1j}^{(\infty)}(0) = 1/(n-1)$ and no transition $v_1 \to v_j$ is preferred over any other. Also, the "lack of information" $H_y^{(\infty)}(0)$ is maximal. As $t \to \infty$, $y_{12}^{(\infty)}(t) \to 1$ and the learning of the transition $v_1 \to v_2$ approaches perfection. At the same time, the "lack of information" $H_y^{(\infty)}(t)$ approaches a minimum.

In the previous section, we showed how the mathematical concept of "information" can be used to improve the accuracy of "learned" predictions. A forthcoming paper will show how to modify the very laws (2)-(4) to provide a comparable improvement in accuracy without directly invoking the entropy function. This modification will illustrate a considerably closer tie between the mathematical concept of "information" and the idea of "learning" herein described.

8. Statistical vs. Deterministic Prediction

Our system of equations can be interpreted as a prediction theory in two different ways. Firstly it is a deterministic prediction theory in which a machine is perturbed by individual inputs and produces individual outputs at specified times. Secondly it is a statistical prediction theory which describes how the probability of given predictions evolves through time. This double interpretation is particularly evident when, for example, $y_{AB}(t) \to 1$ as $t \to \infty$ in a $G^{(\infty)}$ outstar with $\theta_j = \delta_{j2}$. Then we can say both that the prediction B given A occurs with probability 1 or that it always occurs upon demand as $t \to \infty$. The former interpretation describes no particular event *in time*, whereas the latter does. This is the main advantage of a deterministic theory.

Because of this double interpretation, our theory has a clear limitation, which can be described using the outstar of Fig. 12, in which the list AB

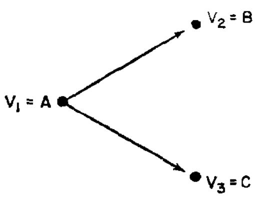

FIG. 12

occurs $\frac{1}{2}$ of the time, the list AC occurs $\frac{1}{2}$ of the time, and the two lists are intermixed evenly in time; i.e., $ABACABACABAC \cdots$ occurs. It is then highly plausible that the averages $(1/t) \int_0^t y_{AB}(v)\, dv$ and $(1/t) \int_0^t y_{AC}(v)\, dv$ come close to $\frac{1}{2}$ as t becomes large (i.e., "statistical" prediction). In particular, it is impossible to predict either B or C with perfect accuracy knowing ACA or ABA has occurred (i.e., "deterministically"). Because of this, the present theory can hope to predict deterministically only lists $ABC \cdots XYZ$ in which no symbol occurs in more than one transition. Otherwise, only statistical prediction is possible.

The modified laws (2)-(4) discussed in the preceding section can be interpreted as a closely related prediction theory which can with some success predict alternating patterns deterministically as well as fulfill the general goal of predicting individual events, in a fixed order, and at prescribed times.

REFERENCES

1. S. GROSSBERG. Some nonlinear networks suggested by learning theory. Thesis, The Rockefeller University, 1967.
2. S. GROSSBERG. Nonlinear difference-differential equations in prediction and learning theory. *Proc. Natl. Acad. Sci.*, **58** (1967), 1329-1334.
3. S. GROSSBERG. Global ratio limit theorems for some nonlinear functional-differential equations. *Bull. Amer. Math. Soc.*, **74** (1968), 95-105.
4. R. G. BUSACKER AND T. L. SAATY. "Finite Graphs and Networks." McGraw-Hill, 1953.

5. L. Cesari. "Asymptotic Behavior and Stability Problems in Ordinary Differential Equations." Academic Press, New York, 1963.
6. L. R. Ford, Jr. and D. R. Fulkerson. "Flows in Networks." Princeton, New Jersey, 1962.
7. A. Halanay. "Differential Equations; Stability, Oscillations, Time Lags." Academic Press, New York, 1966.
8. A. I. Khinchin. "Mathematical Foundations of Information Theory." Dover, New York, 1957.
9. C. E. Osgood. "Method and Theory in Experimental Psychology." Oxford, London, 1953.