INFORMATION AND COMPUTATION 89, 47–94 (1990)

Creating Order in Sequence Spaces with Simple Machines

RUDOLF AHLSWEDE, JIAN-PING YE, AND ZHEN ZHANG

Fakultät für Mathematik, Universität Bielefeld, Universitätsstrasse 1, 4800 Bielefeld 1, West Germany

We intend to open a new research field towards, say, a theory of "creating order" under various constraints. As a prototype of problems guiding our investigations we study models involving sequence spaces. By "creating order" or equivalently "organization" we mean reducing in size the range of outputs by an "organizer" via a permuting channel (a simple machine), when it is fed by a given domain of inputs. The "creation of order" is assumed to come only from the permutation operation in these channels. Four types of "order creation" are considered depending on the structure of the knowledge of the organizer (limitations on mind) about the future input and past output sequences and the kinds of admissible permutations inside the channel (limitations on matter). In any case the organizer's goal is to produce output spaces of minimal cardinality (optimal organization). We present some strategies of ordering and some first and seemingly basic optimality results. After this more technical part of the paper we present some ideas about a general theory of ordering.

INTRODUCTION

A short reflection shows that people spend a large amount of time creating order in various circumstances. We mention a few.

Our homes are daily to be taken care of, we must clean our clothes and even ourselves. Our houses and cars are to be repaired. Garbage must be collected. Rules for human relations are to be set up and violations of laws in a society are to be controlled by the police. Politicians try to improve the organization of a state and relations among countries. Bookkeeping and organization of files constitute a great part of administrative activities. Even the scientists' goal of understanding some aspects of the world can be viewed as an attempt to organize phenomena by some principles.

Our general aim is to start or to contribute to a theory of ordering. In particular we try to understand how much "order" can be created in a "system" under constraints on our "knowledge about the system" and on the "actions we can perform in the system." In this generality it would be permature even to try to give these terms a precise meaning.

0890-5401/90 $3.00

643/89/1-4

At the end of the paper we present illustrative examples and discuss the motivation, thoughts, and "philosophy" guiding our work. Several directions of research are sketched.

In the main body of the paper we restrict ourselves to models involving sequence spaces. Several practical processes, for instance, can be modelled by a sequence of independent identically distributed random variables $(X_t)_{t=1}^n$ with values in a finite set $\mathcal{X}$. Its elements are *physical* objects (such as economical goods and documents) which we want to maintain.

However, we may be interested in rearranging $X_1, ..., X_n$ into a sequence $Y_1, ..., Y_n$ meeting specified goals. For instance, we may wish to achieve $Y_1 \geqslant \cdots \geqslant Y_n$, where "$\geqslant$" is a linear order on $\mathcal{X}$, or we may want to reduce entropy. Typically there are limitations on the capability to create order such as *limitations on matter* in such a way that, stepwise, we can perform only pairwise comparisons, and *limitations on mind* to the extent that while performing an operation we have only partial knowledge of $X^n = X_1, ..., X_n$.

We formulate and investigate models which are motivated by our work on permuting channels (Ahlswede and Kaspi, 1987). Thus we are already confronted with numerous rather interesting mathematical problems. In some cases we have found solutions with surprising answers. They are the first seemingly basic results in this area.

A. *Our Non-probabilistic Model*

Suppose we have a box that contains β objects at time t. We assume that the objects are labelled with numbers from $\mathcal{X} = \{1, 2, ..., \alpha\}$. For simplicity we say "an object i" instead of "an object labelled by i." Thus the content or "state" of the box can be described by a multi-set $s_t = (s_t(1), ..., s_t(\alpha))$, where $s_t(i)$ is the number of i's in the box at time t and $\sum_{i=1}^{\alpha} s_t(i) = \beta$.

Assume now that an *arbitrary* n-length sequence, say $x^n = (x_1, x_2, ..., x_n) \in \mathcal{X}^n$, enters the box iteratively. At time t, x_t enters just after a person $\mathcal{O}$, called the organizer, has thrown out object y^t. Consequently, the state s_t changes to s_{t+1}. s_1 is the initial state and s_{n+1} is the terminal state. We call x^n an input and $y^n = (y_1, ..., y_n)$ an output sequence. The organizer's behaviour must obey the following rules.

Constraints on matter. The organizer can output only objects which he has in the box. At each time t $(1 \leqslant t \leqslant n)$ he must output exactly one object.

Constraints on mind. The organizer's behaviour (strategy) depends on

(a) his knowledge about the time t. The cases where $\mathcal{O}$ has a timer and has no timer are denoted by T^+ and T^-, respectively.

(b) his knowledge about the content of the box. We indicate the situation where $\mathcal{O}$ knows at time t only the state $s_t \in \mathcal{S}$, the set of all states, by

O . If in addition he knows the natural order of the objects in the box, that is, the order according to their entrance times, we denote this by O^+.

(c) his knowledge about x^n and his past actions. We assume this to be the following nature: At time t (regardless of whether we are in case T^+ or T^-) with state of the box s_t the organizer can see the incoming letters $x_t, x_{t+1}, \ldots, x_{t+\varphi}$ and he remembers (or can see) the output letters $y_{t-\pi}, y_{t-\pi+1}, \ldots, y_{t-1}$ when he outputs y_t. With this understanding we describe the memory by a triple (π, β, φ). Here φ measures the time the organizer can forsee, β is the number of objects in the box, which we also call working area, and π measures the past time for which the organizer has a memory. Loosely speaking π, β, and φ represent the past, present, and future "memory" of the system.

Note that input and output sequences are always ordered. Concerning timer and order in the box we have the four possibilities (T^-, O^-), (T^+, O^-), (T^-, O^+), and (T^+, O^+).

If such a pair, for instance (T^-, O^-), and also (π, β, φ) are specified, then for every n we have a set of strategies $\mathscr{F}_n(\pi, \beta, \varphi; T^-, O^-)$ which are based on the knowledge available to $\mathcal{O}$. Every strategy $f_n: \mathscr{X}^n \times \mathscr{S} \to \mathscr{X}^n$ in $\mathscr{F}_n(\pi, \beta, \varphi; T^-, O^-)$ assigns to a pair (x^n, s_1) an output sequence $y^n = f_n(x^n, s_1)$. We denote by $\mathscr{Y}(f_n)$ the image of $\mathscr{X}^n \times \mathscr{S}$ under f_n. This is the set of output sequences which can occur in the worst case. Let $\|\mathscr{Y}(f_n)\|$ stand for the cardinality of $\mathscr{Y}(f_n)$.

We are now prepared to introduce the basic performance criteria:

Size and rate. Define the size

$$N_\alpha^n(\pi, \beta, \varphi) = \min\{\|\mathscr{Y}(f_n)\| : f_n \in \mathscr{F}_n(\pi, \beta, \varphi; T^-, O^-)\} \tag{1.1}$$

and the rate

$$v_\alpha(\pi, \beta, \varphi) = \lim_{n \to \infty} \frac{1}{n} \log N_\alpha^n(\pi, \beta, \varphi). \tag{1.2}$$

Here the letters N and v have been chosen to indicate that we have no timer and no order in the box. Analogously, we define in the case (T^-, O^+) the quantities $O_\alpha^n(\pi, \beta, \varphi)$, $\omega_\alpha(\pi, \beta, \varphi)$ and in the case (T^+, O^-) the quantities $T_\alpha^n(\pi, \beta, \varphi)$, $\tau_\alpha(\pi, \beta, \varphi)$. Finally, in the case (T^+, O^+) we write $G_\alpha^n(\pi, \beta, \varphi)$ and $\gamma_\alpha(\pi, \beta, \varphi)$. *Except in Sections* 5 *and* 8 *we assume throughout this paper that the alphabets for the input and output spaces are binary*, that is, $\mathscr{X} = \mathscr{Y} = \{0, 1\}$ and $\alpha = 2$. Therefore, in most of these quantities we omit the index α, if $\alpha = 2$.

B. *Active Memory*

In the model just introduced the memory may be termed passive, because $\mathcal{C}$ simply collects certain data about the future and past. Instead or in addition there is now storage space of size m attached to the box, where m bits of information can be stored. $\mathcal{C}$ is free to store there any information he has at any time and also to destroy a part of this information in order to have space for new information relevant to him.

We are thus led to study quantities $N_{\Sigma}^{n}(\pi, \beta, \varphi, m)$, etc.

C. *A Probabilistic Model*

Suppose now that the input sequence is an i.i.d. sequence of RVs $(X_t)_{t=1}^{n}$. Also, the initial content of the box may be produced by such a sequence. The constraints on matter and mind from A are again meaningful. If $\mathcal{O}$ follows strategy f, then this gives rise to an output process $(Y_t)_{t=1}^{n}$. The performance of f is now measured by the entropy $H(Y^n)$ resp. the mean entropy $\overline{H}_f = \overline{\lim}_{n\to\infty} (1/n)\, H(Y^n)$.

The goal now is to minimize this quantity. There may also be active memory. Several rather difficult problems arise; a beginning of their analysis and of active memory has been made in work by Ahlswede and Zhang (1989).

This probabilistic model and our non-probabilistic model are extremal cases of the more general model characterized by specifying sets of probability distributions $\mathcal{P}_n$ on $\mathcal{X}^n$ for every n. The understanding here is that the distribution $P_n \in \mathcal{P}_n$ governing the input selection is unknown to $\mathcal{O}$.

The paper is organized as follows: All of our results are for the non-probabilistic model with passive memory. They are presented and proved in Sections 2 to 6. The proofs are often based on new combinatorial results of some independent interest. Difference equations play a key role. In Section 7 our most basic results are surveyed in a chart and in Section 8 we state several conjectures. Many problems are left unsolved and may challenge other mathematicians to work on them. Finally, in Section 9 we contribute several models and ideas for the theory of ordering. This opens a new area of research.

2. Formulas for $N^n(\pi, \beta, \varphi)$ in Basic Extremal Cases

We begin our study of the functions $N_{\alpha}^{n}(\pi, \beta, \varphi)$, defined in (1.1), for binary alphabet $\mathcal{X} = \{0, 1\}$. Here the set of all states $\mathcal{S}$ can be identified with the set $S = \{s: 0 \leqslant s \leqslant \beta\}$, where s counts the number of 1's in the box.

Even in this case it is very difficult to find a formula for all values of π and φ. However, for basic extremal cases of these values we have found solutions.

We begin with the case in which the organizer $\mathcal{O}$ has no knowledge about past and future.

THEOREM 1. *$N^n(0, \beta, 0) = 2^n$ for all $\beta \geqslant 1, n \geqslant 1$.*

Proof. Since at any time t, $1 \leqslant t \leqslant n$, $\mathcal{C}$ knows solely the state s_t, but not t, his strategy f_n is already determined by a function $f: S \to \mathcal{X}$, with the property

$$f(0) = 0, \qquad f(\beta) = 1. \tag{2.1}$$

We proceed by induction on (β, n). For the induction beginning we have $N^n(0, 1, 0) = 2^n$ and $N^1(0, \beta, 0) = 2$.

Now we distinguish between two cases.

Case 1. There is an s', $1 \leqslant s' \leqslant \beta - 2$, with

$$f(s') = 1 \qquad \text{and} \qquad f(s'+1) = 0.$$

Partition S into the two sets $\{0, 1, ..., s'\}$ and $\{s'+1, ..., \beta\}$. By our assumptions on f these state sets are left invariant while we follow f_n. Therefore, if $M^n(f, \beta)$ counts the number of output sequences under f_n

$$M^n(f, \beta) \geqslant \max(N^n(0, s', 0), \qquad N^n(0, \beta - s' - 1, 0)) \geqslant 2^n \tag{2.2}$$

by the induction hypothesis.

Case 2. There is an s^*, $0 \leqslant s^* \leqslant \beta - 1$, such that for the sets $S_0 = \{0, 1, ..., s^*\}$, $S_1 = \{s^*+1, ..., \beta\}$

$$f(s) = \begin{cases} 0 & \text{for} \quad s \in S_0 \\ 1 & \text{for} \quad s \in S_1. \end{cases}$$

Note that in the first step under f_n, S_0 is transformed into the invariant set $S_0 \cup \{s^*+1\}$ and S_1 is transformed into the invariant set $S_1 \cup \{s^*\}$. Therefore

$$M^n(f, \beta) = M^{n-1}(f, s^*+1) + M^{n-1}(f, \beta - s^*) \geqslant 2 \cdot 2^{n-1} \tag{2.3}$$

by the induction hypothesis. Since obviously, $M^n(f, \beta) \leqslant |\mathcal{X}^n| = 2^n$, the result follows.

Actually we have proved that for every strategy f_n, $M^n(f, \beta) = 2^n$. The cases $\alpha > 2$ are much more difficult to analyse. Partial results can be found in Section 7.

Next we consider the case of complete knowledge about past and future.

THEOREM 2. (i) $N^n(\infty, \beta, \infty) = 2^{\lceil n/\beta \rceil}$ *for* $n \geq 1$.

(ii) $v(\infty, \beta, \infty) = 1/\beta$.

Proof. Because (ii) is an immediate consequence of (i), we must prove (i). The key idea consists in introducing a set $\mathscr{D}(n, \beta)$, which we now define.

Let $\delta_0 = (0, ..., 0)$ and $\delta_1 = (1, ..., 1)$ be sequences of length β. Furthermore, for $r = n \bmod \beta$ we define the sequences $\delta_0 = (0, ..., 0)$, $\delta_1 = (1, ..., 1)$ of lengths r. Set $v = \lfloor n/\beta \rfloor$.

$$\mathscr{D}(n, \beta) = \{x(\delta) \in \mathscr{X}^n : x(\delta) = (\delta_{(1)}, ..., \delta_{(v)}, \delta_{(v+1)}) \text{ with } \delta_{(v+1)} \in \{\delta_0, \delta_1\} \text{ and } \delta_{(i)} \in \{\delta_0, \delta_1\} \text{ for } i = 1, ..., v\}. \quad (2.4)$$

Note that

$$|\mathscr{D}(n, \beta)| = 2^{\lceil n/\beta \rceil}. \quad (2.5)$$

We prove the following two facts:

Every $x^n \in \mathscr{X}^n$ can be encoded into an $x(\delta) \in \mathscr{D}(n, \beta)$. (2.6)

No two $x(\delta)$, $x(\delta') \in \mathscr{D}(n, \beta)$ with $\delta_{(i)} \neq \delta'_{(i)}$ for an $i \leq \lceil n/\beta \rceil - 1$ can be encoded into the same y^n, if the initial states are equal. (2.7)

Ad (2.6). At times $\beta i + 1$ $(i = 0, 1, 2, ...)$ we look at $s_{\beta i+1} + |x_{\beta i+1}, ..., x_{\beta(i+1)-1}|_1$, that is, the number of 1's in the box and among the $\beta - 1$ incoming letters. If this number is not smaller than β, then we send a 1 and continue sending 1's β times. Note that this is possible because at the τth, $\tau \leq \beta$, step we have

$$s_{\beta i+\tau} + |x_{\beta i+\tau}, ..., x_{\beta(i+1)-1}|_1 \geq \beta - (\tau - 1)$$

and therefore $s_{\beta i+\tau} \geq \beta - (\tau - 1) - (\beta - \tau) \geq 1$, and because observation of 1-past already tells us that we are in the process of sending β 1's.

Otherwise we have at least β 0's among the initial two $\beta - 1$ letters and we send β successive 0's. Starting with $i = 0$ we repeat this until $i = \lfloor n/\beta \rfloor - 1$. If now $r = 0$, then we are done, and otherwise we have the situation $i = \lfloor n/\beta \rfloor$. Here we proceed essentially as before; the only difference is that now we send r 1's, if there are at least $r + 1$ 1's among the letters $x_{\lfloor n/\beta \rfloor \beta + 1}, ..., x_n$ and in the box. Otherwise, since $\beta \geq r + 1$, we send r 0's.

Ad (2.7). Let i be minimal with $\delta(i) \neq \delta'_{(i)}$. If the outputs are the same from time 1 to time βi, then at time $\beta i + 1$ there must be β 1's in the box

for one of $x(\delta)$ and $x(\delta')$ and β 0's for the other one. Thus necessarily $y_{\beta i+1} \neq y'_{\beta i+1}$ and $y^n \neq y'^n$, as claimed. Now (2.7) and (2.5) imply

$$N^n(\infty, \beta, \infty) \geqslant |\mathscr{D}(n, \beta)| = 2^{\lceil n/\beta \rceil},$$

and by (2.6), $N^n(\infty, \beta, \infty) \leqslant |\mathscr{D}(n, \beta)|$. The proof is complete.

Remark. In our (optimal) strategy yielding (2.6) we have used only

(a) knowledge of time,

(b) knowledge of the future for $\varphi = \beta - 1$,

(c) knowledge of the past for $\pi = 1$.

Here (a) follows from the knowledge of the ∞-past or the ∞-future. Furthermore, N^n, T^n, etc., are obviously monotonically decreasing in π, β, and φ. Thus we have established the following generalization of Theorem 2.

$$T^n(\pi, \beta, \varphi) = N^n(\pi, \beta, \varphi) = 2^{\lceil n/\beta \rceil} \quad \text{for} \quad \varphi \geqslant \beta - 1 \quad \text{and} \quad \pi \geqslant 1. \tag{2.8}$$

Calculations show that this equation does not hold for $\pi = 0$. It is therefore very remarkable that in the case in which $\mathscr{O}$ has knowledge only about the ∞-future the optimal *rate* is still $1/\beta$. We now describe a strategy which achieves this bound.

For a sequence $a^m = (a_1, ..., a_m)$ we write a_ℓ^m for $(a_\ell, ..., a_m)$. At time $t = 1$ the encoder knows that $z_1^n = (x_1^n, s_1)$, where s_1 describes the state of the box and x_1^n describes the future. Generally at time t, $z_t^n = (x_t^n, s_t)$ describes the knowledge of the encoder. We also use for $i < t$ the notation $z_t^i = s_t$, which indicates the knowledge about the state of the box.

Further, let $\langle z_t^i \mid \varepsilon \rangle$ count at time t how often ε occurs in the box and in x_t^i, if $i \geqslant t$.

For a positive integer ℓ we denote by $\ell \bmod^* \beta$ the number p for which there exists a q with

$$\ell = q\beta + p, \qquad 1 \leqslant p \leqslant \beta. \tag{2.9}$$

For the definition of our encoding procedure ψ we distinguish among four cases.

If $\langle z_t^n \mid \varepsilon \rangle = 0$, then we speak of the ε-simple case. Here we define $\psi(z_t^n) = 1 - \varepsilon$ and $\psi(z_s^n) = 1 - \varepsilon$ for $t \leqslant s \leqslant n$ is our only choice.

For the description of ψ in *all other cases* we use

$$a_t = \langle z_t^n \mid 0 \rangle \bmod^* \beta, \qquad b_t = \langle z_t^n \mid 1 \rangle \bmod^* \beta. \tag{2.10}$$

These cases are called regular, if $a_t + b_t > \beta + 1$; critical, if $a_t + b_t = \beta + 1$;

and ambiguous, if $a_t + b_t < \beta + 1$. In regular and critical cases ψ is defined by

$$\psi(z_t^n) = \begin{cases} 0 & \text{if } \langle z_t^{a_t + b_t + t - \beta - 2} \mid 0 \rangle \geq a_t \\ 1 & \text{if } \langle z_t^{a_t + b_t + t - \beta - 2} \mid 1 \rangle \geq b_t \end{cases} \tag{2.11}$$

and in ambiguous cases ψ always takes the value 1. ψ is well-defined in the regular case, because

$$\begin{aligned} &\langle z_t^{a_t + b_t + t - \beta - 2} \mid 0 \rangle + \langle z_t^{a_t + b_t + t - \beta - 2} \mid 1 \rangle \\ &\quad = a_t + b_t + t - \beta - 2 + \beta - t + 1 = a_t + b_t - 1 \end{aligned}$$

and exactly one of the alternatives in (2.11) is true. In the critical case the equation $a_t + b_t = \beta + 1$ implies that exactly one of the alternatives

$$\beta - b_t \geq a_t, \qquad s_t \geq b_t \tag{2.12}$$

holds.

In the ambiguous case at least one of these alternatives holds, but also both may hold. Our convention with ψ is always to choose the second. We show now that ψ can be implemented; that is, an object (0 or 1) prescribed by it is always in the box.

If the simple case arises, we are done. In any other case let us assume, for example, that the second alternative is true. Then from

$$s_t = \begin{cases} \langle z_t^{a_t + b_t + t - \beta - 2} \mid 1 \rangle & \text{in critical or ambiguous cases} \\ \langle z_t^{a_t + b_t + t - \beta - 2} \mid 1 \rangle - \langle x_t^{a_t + b_t + t - \beta - 2} \mid 1 \rangle & \text{in regular cases} \end{cases}$$

we conclude that

$$\begin{aligned} &s_t \geq b_t \geq 1 \qquad \text{in critical or ambiguous cases} \\ &s_t \geq b_t - (a_t + b_t + t - \beta - 2 - t + 1) = \beta + 1 - a_t \\ &\quad \geq \beta + 1 - \beta = 1 \qquad \text{in regular cases.} \end{aligned}$$

Similarly, if the first alternative holds, it is also possible to send a 0.

We now analyze ψ.

The simple cases are settled. We can assume henceforth that $1 \leq a_t, b_t \leq \beta$. Suppose first that at time t, ψ takes the value 0, that is,

$$\langle z_t^{a_t + b_t + t - \beta - 2} \mid 0 \rangle \geq a_t.$$

Then either $a_t = 1$ or $a_t \geq 2$.

In the first case $a_t + b_t = 1 + b_t \leq \beta + 1$ and since $\psi = 0$ we are necessarily in the critical case with $b_t = \beta$. At time $t+1$ we obtain $a_{t+1} = \beta$, $b_{t+1} = b_t = \beta$ and therefore $a_{t+1}, b_{t+1} \geq 2$ (as in the second case).

In the second case, however, $a_{t+1} = a_t - 1$, $b_{t+1} = b_t$, and

$$\langle z_{t+1}^{a_{t+1}+b_{t+1}+t+1-\beta-2} \mid 0\rangle$$

$$= \langle z_{t+1}^{a_t+b_t+t-\beta-2} \mid 0\rangle$$

$$= \begin{cases} \beta - s_{t+1} & \text{if } a_t + b_t \leqslant \beta + 2 \\ \beta - s_{t+1} + \langle x_{t+1}^{a_t+b_t+t-\beta-2} \mid 0\rangle & \text{otherwise} \end{cases}$$

$$= \begin{cases} \beta - s_t - 1 + \langle x_t \mid 0\rangle & \text{if } a_t + b_t \leqslant \beta + 2 \\ \beta - s_t - 1 + \langle x_t \mid 0\rangle + \langle x_{t+1}^{a_t+b_t+t-\beta-2} \mid 0\rangle & \text{otherwise} \end{cases}$$

$$= \begin{cases} \langle z_t^{a_t+b_t+t-\beta-2} \mid 0\rangle - 1 + \langle x_t \mid 0\rangle & \text{if } a_t + b_t < \beta + 2 \\ \langle z_t^{a_t+b_t+t-\beta-2} \mid 0\rangle - 1 & \text{if } a_t + b_t = \beta + 2 \\ \langle z_t^{a_t+b_t+t-\beta-2} \mid 0\rangle - 1 & \text{if } a_t + b_t > \beta + 2. \end{cases}$$

In any case we have

$$\langle z_{t+1}^{a_t+b_t+t-\beta-2} \mid 0\rangle \geqslant a_t - 1 = a_{t+1}$$

and we continue to send 0 as long as we are in the regular or the critical case.

As soon as we enter the ambiguous case we continue to send 1's. Again the foregoing arguments apply. Now we can send b_t times a 1. Moreover, afterwards we are either in the simple case or $b_{t+b_t} = \beta$. Therefore, from now on, if we can send 1, then we will always send it in blocks of length β. The same scheme applies to the sending of 0 except when we enter the ambiguous case. However, after we have sent 1's once in blocks of length β ending at s, say, then $a_s + b_s = a_s + \beta \geqslant \beta + 1$; that is, if it is possible to send 0, then the ambiguous case cannot occur while we are sending a_s 0's, because the b's retain the value β. But now $a_{s+a_s} = \beta$, if we do not enter the simple case, and from now on 0's too, are sent in blocks of length β.

Therefore the number of possible output sequences does not exceed

> (number of possible 0-strings of length $\leqslant \beta$) $\times$ (number of possible 1-strings of length $\leqslant \beta$) $\times 2^{\lceil n/\beta \rceil} \times$ (number of possible lengths of the last string) + (number of possible 1-strings of length $\leqslant \beta$) $\cdot 2^{\lceil n/\beta \rceil} \cdot$ (number of possible lengths of the last string) $\leqslant (\beta^3 + \beta^2)\, 2^{\lceil n/\beta \rceil} \leqslant \beta^3(1 + 1/\beta)\, 2^{\lceil n/\beta \rceil} \leqslant \frac{3}{2}\beta^3 2^{\lceil n/\beta \rceil}$.

We summarize our findings.

THEOREM 2*. (i) $N^n(\infty, \beta, \varphi) = N^n(\infty, \beta, \infty) = 2^{\lceil n/\beta \rceil}$ *for* $\varphi \geqslant \beta - 1$.

(ii) $T^n(\pi, \beta, \varphi) = N^n(\pi, \beta, \infty) = 2^{\lceil n/\beta \rceil}$ *for* $\varphi \geqslant \beta - 1$ *and* $\pi \geqslant 1$.

(iii) $v(0, \beta, \infty) = v(\infty, \beta, \infty) = 1/\beta$.

3. Infinite Past

It is remarkable that v does not depend on the past, if we know the ∞-future. On the other hand we shall see below that knowledge of the future does help, if we know the ∞-past. These cases are more difficult to analyse. *Roughly* speaking we have $N^n(\infty, \beta, 0) \sim_v N^n(0, \beta, \infty)$ and in this sense knowledge of the future is more valuable than knowledge of the past. Again we can use the knowledge of time. Moreover, we shall prove that in this case also $N^n(\infty, \beta, 0) = O^n(\infty, \beta, 0)$.

A. ∞-Past and O-Future

The following concept of a weighted tree turns out to be essential.

Definition. Let B be a binary tree with the properties:

(i) The weight of an edge is a number in $\{1, \ldots, \beta\}$.

(ii) These numbers add to at most $\beta + 1$ for the two edges leaving an internal node.

(iii) Every path from the root to a terminal node has length n, if edges are counted with their weight.

(iv) The two edges leaving an internal node are labelled with 0 and 1.

Let $\mathscr{B}(n, \beta)$ denote the set of all those trees and let $C(B)$ be the number of terminal nodes in B.

An important quantity is

$$C(n, \beta) = \min_{B \in \mathscr{B}(n, \beta)} C(B). \tag{3.1}$$

With every tree $B \in \mathscr{B}(n, \beta)$ we can associate a strategy $g(B)$ as follows:

If s_1 is the state of the box at the time $t = 1$ and l_ε is the length of the edge leaving the root of B and labelled by $\varepsilon \in \{0, 1\}$, send l_1 1's if $s_1 \geq l_1$, and otherwise send l_0 0's. Since $l_0 + l_1 \leq \beta + 1$, this is possible. We reach a new node of the tree and a new state afterwards. Now just iterate the procedure until $t = n$ and at the same time a terminal node is reached.

The number of possible output sequences does not exceed $C(B)$. This is our first result.

Proposition 1. $N^n(\infty, \beta, 0) \leq C(n, \beta)$.

Via several lemmas, we next prove

Proposition 2. $O^n(\infty, \beta, 0) = N^n(\infty, \beta, 0) \geq C(n, \beta)$. *Then we shall evaluate $C(n, \beta)$ in Proposition 3, below. The consequences of these propositions are stated in Theorem 3.*

The Lower Bound. By the definitions $N^n(\infty, \beta, 0) \geq O^n(\infty, \beta, 0)$ and while analysing any strategy f in the case O , we show also that actually equality holds.

For output y^i define

$$\mathscr{Y}(f, y^i) = \{ y^n \in \mathscr{Y}(f) \colon y^n = (y^i, y^n_{i+1}) \} \tag{3.2}$$

$$\mathscr{S}(f, y^i) = \{ s : s \text{ occurs as the state at time } i+1 \text{ for output } y^i \}. \tag{3.3}$$

An optimal strategy minimizes $|\mathscr{Y}(f, y^i)|$ for every y^i, because the infinite past, and thus y^i, is known. However, for this minimization only the knowledge of $\mathscr{S}(f, y^i)$ is relevant; that is, the actual value of y^i does not matter. For any optimal strategy under consideration we can therefore write the quantities in (3.3), (3.2) as $\mathscr{S}(y^i)$ and $\mathscr{Y}^m(\mathscr{S}(y^i))$, $m = n - i$. Their analysis reduces to the following:

> Given $\mathscr{S} \subset \{0, 1, ..., \beta\}$ as the set of possible states to start with, how can we lower bound $|\mathscr{Y}(\mathscr{S})|$?

Here it is understood that $m = n - i$ steps are to be taken. Let $\mathscr{S}_0$ (resp. $\mathscr{S}_1$) be the subset of $\mathscr{S}$ for which the strategy sends 0 (resp. 1). If the strategy may depend on the order (O^+), both 0 and 1 can be sent for the same state s. Therefore $\mathscr{S}_0$ and $\mathscr{S}_1$ need not be disjoint. Of course

$$\mathscr{S}_0 \cup \mathscr{S}_1 = \mathscr{S}. \tag{3.4}$$

However, $\mathscr{Y}(\mathscr{S})$ will only be decreased by sending 0 (or 1) for all $s \in \mathscr{S}_0 \cap \mathscr{S}_1$. Thus $N^n(\infty, \beta, 0) = O^n(\infty, \beta, 0)$ holds and we can always assume that

$$\mathscr{S}_0 \cap \mathscr{S}_1 = \varnothing. \tag{3.5}$$

The following formulas follow from the fact that after a letter is sent out of the box, both 0 and 1 can enter the box.

$$\mathscr{S}_1(1) = \bigcup_{s \in \mathscr{S}_1} \{s-1, s\}, \qquad \mathscr{S}_0(0) = \bigcup_{s \in \mathscr{S}_0} \{s, s+1\}. \tag{3.6}$$

A first simple observation is

LEMMA 1. (a) *If* $\mathscr{S}' \subset \mathscr{S}$, *then* $|\mathscr{Y}(\mathscr{S}')| \leq |\mathscr{Y}(\mathscr{S})|$.

(b) *If* $\mathscr{S}' \cup \mathscr{S}'' \supset \mathscr{S}$, *then* $|\mathscr{Y}(\mathscr{S}')| + |\mathscr{Y}(\mathscr{S}'')| \geq |\mathscr{Y}(\mathscr{S})|$.

Proof. (a) Just use the strategy for $\mathscr{S}'$ which is induced by an optimal strategy for $\mathscr{S}$, that is, define

$$\mathscr{S}'_1 = \mathscr{S}' \cap \mathscr{S}_1 \qquad \text{and} \qquad \mathscr{S}'_0 = \mathscr{S}' \cap \mathscr{S}_0.$$

(b) By (a) it suffices to consider the case

$$\mathscr{S}' \cap \mathscr{S}'' = \varnothing, \qquad \mathscr{S}' \cup \mathscr{S}'' = \mathscr{S}.$$

Here strategies f', f'' for $\mathscr{S}'$, $\mathscr{S}''$ induce a strategy f for $\mathscr{S}$ as follows:

$$\mathscr{S}_1 = \mathscr{S}'_1 \cup \mathscr{S}''_1, \qquad \mathscr{S}'_0 = \mathscr{S}'_0 \cup \mathscr{S}''_0.$$

Thus in obvious notation

$$\mathscr{Y}(\mathscr{S}', f') \cup \mathscr{Y}(\mathscr{S}'', f'') \supset \mathscr{Y}(\mathscr{S}, f),$$

which implies (b).

Our key auxiliary result is

LEMMA 2. (a) *If* $0 \notin \mathscr{S}$ (resp. $\beta \notin \mathscr{S}$), *then sending* 1 (*resp.* 0) *for* all *states in* $\mathscr{S}$ *is optimal.*

(b) *If* $\mathscr{S} = \bar{\mathscr{S}} + c$, *where the bar denotes complementation,* c *is an integer and the addition is that for integers, then for an optimal strategy*

$$|\mathscr{Y}(\mathscr{S})| = |\mathscr{Y}(\bar{\mathscr{S}})|.$$

Proof. We proceed by induction in $m = n - i$. For $i = n$ no further letter is sent and the statements are vacuously true.

$n - i - 1 \to n - i$. (a) If $0 \notin \mathscr{S}$, then $\mathscr{S} - 1 \subset \{0, 1, ..., \beta\}$ is defined and thus $\mathscr{S}_0(0) - 1$ is also defined. By the induction hypothesis for (b) therefore

$$|\mathscr{Y}(\mathscr{S}_0(0))| = |\mathscr{Y}(\mathscr{S}_0(0) - 1)|. \tag{3.7}$$

Furthermore, for our strategy

$$\mathscr{S}(1) = \bigcup_{s \in \mathscr{S}} \{s - 1, s\}, \qquad 0 \notin \mathscr{S} \tag{3.8}$$

and thus by (3.6) for the optimal strategy

$$\begin{aligned}\mathscr{S}(1) &= \Big(\bigcup_{s \in \mathscr{S}_1} \{s - 1, s\} \Big) \cup \Big(\bigcup_{s \in \mathscr{S}_0} \{s - 1, s\} \Big) \\ &= \mathscr{S}_1(1) \cup (\mathscr{S}_0(0) - 1).\end{aligned}$$

Application of Lemma 1 gives

$$|\mathscr{Y}(\mathscr{S}(1))| \leqslant |\mathscr{Y}(\mathscr{S}_1(1))| + |\mathscr{Y}(\mathscr{S}_0(0)-1)|$$

and thus by (6.7)

$$|\mathscr{Y}(\mathscr{S}(1))| \leqslant |\mathscr{Y}(\mathscr{S}_1(1))| + |\mathscr{Y}(\mathscr{S}_0(0))|.$$

Since the case $\beta \notin \mathscr{S}$ is symmetrically the same, (a) is proved under the induction hypothesis for (b).

(b) For $c=0$ nothing is to be proved and for the case $c \neq 0$, (a) applies to both set $\mathscr{S}$ and set $\bar{\mathscr{S}}$, and the minimal values $|\mathscr{Y}(\mathscr{S})|$, $|\mathscr{Y}(\bar{\mathscr{S}})|$ are assumed for strategies described in (a).

Following these strategies, after one step we have two sets of states, $\mathscr{S}(y_{i+1})$ and $\bar{\mathscr{S}}(\bar{y}_{i+1})$.

Since

$$\mathscr{S}(y_i+1) = \bigcup_{s \in \mathscr{S}} \{s - y_{i+1}, s+1-y_{i+1}\}$$

and

$$\bar{\mathscr{S}}(\bar{y}_{i+1}) = \bigcup_{s \in \mathscr{S}-c} \{s - \bar{y}_{i+1}, s+1-\bar{y}_{i+1}\},$$

there exists a $d \in \{c, c+1, c-1\}$ with

$$\mathscr{S}(y_{i+1}) = \bar{\mathscr{S}}(\bar{y}_{i+1}) + d.$$

By the induction hypothesis for (b) therefore

$$|\mathscr{Y}(\mathscr{S}(y_{i+1}))| = |\mathscr{Y}(\bar{\mathscr{S}}(\bar{y}_{i+1}))|.$$

Since

$$\mathscr{Y}(\mathscr{S}) = y_{i+1} * \mathscr{Y}(\mathscr{S}(y_{i+1}))$$

and

$$\mathscr{Y}(\bar{\mathscr{S}}) = \bar{y}_{i+1} * \mathscr{Y}(\bar{\mathscr{S}}(y_{i+1})),$$

the result follows.

Using Lemmas 1 and 2 we can now derive another basic auxiliary result.

LEMMA 3. *For any* $\mathscr{S} \subset \{0, 1, ..., \beta\}$, $\mathscr{S} \neq \varnothing$,

$$|\mathscr{Y}(\mathscr{S})| \geqslant |\mathscr{Y}(\{0, 1, ..., |\mathscr{S}|-1\})|.$$

Proof. Let $\{(m, \tau): 1 \leqslant m \leqslant n; 1 \leqslant \tau \leqslant \beta + 1\}$ be ordered lexicographically, that is, $(m, \tau) < (m', \tau')$ iff $m < m'$ or $m = m'$ and $\tau < \tau'$. τ stands for $|\mathscr{S}|$. We proceed by induction in this well-ordered set.

$(m, \tau) = (1, 1)$. If s is the one state, apply Lemma 2(b) with $c = -s$.

Induction. Assume the truth of the inequality for $(m', \tau') < (m, \tau)$. Let an optimal strategy achieving $|\mathscr{Y}(\mathscr{S})|$ send 0 for states in $\mathscr{S}_0$ and 1 for states in $\mathscr{S}_1$. If $\tau = \beta + 1$ nothing is to be proved and otherwise we can assume by Lemma 2(a) that $\mathscr{S}_1 = \varnothing$ or $\mathscr{S}_0 = \varnothing$. Furthermore, it suffices to consider the case $\mathscr{S}_1 = \varnothing$, because by Lemma 2(b) the case $\mathscr{S}_0 = \varnothing$ is symmetrically the same. Again by Lemma 2(a) we know that it is optimal always to send 0 for the set of states $\{0, 1, ..., |\mathscr{S}| - 1\}$. Therefore

$$|\mathscr{Y}(\mathscr{S})| = |0 * \mathscr{Y}^{m-1}(\mathscr{S} \cup \mathscr{S} + 1)|, \tag{3.9}$$

$$|\mathscr{Y}(\{0, 1, ..., |\mathscr{S}| - 1\})| = |0 * \mathscr{Y}^{m-1}(\{0, 1, ..., |\mathscr{S}|\})|. \tag{3.10}$$

Since $|\mathscr{S} \cup \mathscr{S} + 1| \geqslant |\mathscr{S}| + 1$, by the induction hypothesis and Lemma 1(a), we know that

$$|\mathscr{Y}^{m-1}(\mathscr{S} \cup \mathscr{S} - 1)| \geqslant |\mathscr{Y}^{m-1}(\{0, 1, ..., (\mathscr{S})\})|$$

and the result follows with (3.9) and (3.10).

Proof of Proposition 2. Lemmas 2 and 3 have the following important consequence. We start with $\mathscr{S} = \{0, 1, ..., \beta\}$. If $\mathscr{S}_1$ and $\mathscr{S}_0$ are the state sets for an optimal strategy, then

$$\begin{aligned} |\mathscr{Y}(\{0, 1, ..., \beta\})| &= |\mathscr{Y}(\mathscr{S}_1)| + |\mathscr{Y}(\mathscr{S}_0)| \\ &\geqslant |\mathscr{Y}(\{0, 1, ..., |\mathscr{S}_0| - 1\}) + |\mathscr{Y}(\{|\mathscr{S}_0|, ..., \beta\})|. \end{aligned}$$

If $\tau = |\mathscr{S}_0|$, then for $s < \tau$ we should always send 0 and for $s \geqslant \tau$ we should send 1. Moreover, by Lemma 2 it is still optimal, if in the case $s < \tau$ we send a 0 $\beta - \tau + 1$ times and in the case $s \geqslant \tau$ we send 1 τ times. Since necessarily $0 \in \mathscr{S}_0$ and $\beta \in \mathscr{S}_1$ we have $1 \leqslant \tau \leqslant \beta$. Thus this optimal strategy corresponds to a tree in the class $\mathscr{B}(n, \beta)$.

Evaluation of $C(n, \beta)$. It seems intuitively clear that trees with minimal C should be as balanced as integral numbers permit; that is, most edge lengths should be close to $\beta/2$.

To obtain an exact and simple formula for $C(n, \beta)$ seems to be somewhat tricky for even β. It involves a minimization over a system of linear difference equations. We confine ourselves here to the determination of the rate of growth.

PROPOSITION 3. $\lim_{n\to\infty}(1/n)\log C(n,\beta)=\log\lambda^*$, *where* λ^* *is the largest positive root of* $\lambda^{\beta+1}=\lambda^{\lceil(\beta+1)/2\rceil}+\lambda^{\lfloor(\beta+1)/2\rfloor}$. *In particular* $\log\lambda^*=2/(\beta+1)$ *for odd* β.

Proof. By definition of $C(n,\beta)$ we have the recurrence relation

$$C(n,\beta)=\min\Bigg\{C(n-l_0,\beta)+C(n-l_1,\beta):1\leqslant l_0,l_1\leqslant\beta,\ \sum_{i=0}^{1}l_i\leqslant\beta+1\Bigg\},\tag{3.11}$$

$$C(0,\beta)=1.$$

We can modify a tree $B\in\mathscr{B}(n,\beta)$ to a tree $B'\in\mathscr{B}(n,\beta)$ by lengthening all edges so that $\sum_{i=0}^{1}l_i=\beta+1$ holds for all internal nodes and then cutting the tree at depth n. Clearly, $C(B')\leqslant C(B)$ and we can therefore write

$$C(n,\beta)=\min_{1\leqslant l\leqslant\beta}C(n-l,\beta)+C(n-\beta-1+l,\beta)\quad\text{for}\quad n\geqslant\left\lceil\frac{\beta+1}{2}\right\rceil.\tag{3.12}$$

An upper bound for $\overline{\lim}_{n\to\infty}(1/n)\log C(n,\beta)$ is readily obtained by determining the rate of growth of $C_n^{(l)}$ satisfying

$$C_n^{(l)}=C_{n-l}^{(l)}+C_{n-\beta-1+l}^{(l)}\quad\text{for}\quad n\geqslant\left\lceil\frac{\beta+1}{2}\right\rceil\tag{3.13}$$

and by minimizing these rates over l. We obtain the characteristic equation

$$\lambda^n=\lambda^{n-l}+\lambda^{n-\beta-1+l},\tag{3.14}$$

which can be written in the form

$$\lambda^{\beta+1}=\lambda^{\beta+1-l}+\lambda^l,\qquad\lambda>1.\tag{3.15}$$

Allowing l to take any real value, we see that

$$f(l)=\lambda^{\beta+1-l}+\lambda^l=\exp\{(\beta+1-l)\log\lambda\}+\exp\{l\log\lambda\}$$

has first and second derivatives

$$f'(l)=\log\lambda(\exp\{l\log\lambda\}-\exp\{(\beta+1-l)\log\lambda\})$$

$$f''(l)=(\log\lambda)^2 f(l)>0.$$

The minimum occurs for $l=(\beta+1)/2$, and by convexity the smallest value

for integers occurs at $l = \lfloor (\beta+1)/2 \rfloor, \lceil (\beta+1)/2 \rceil$. We have thus seen that for all $\lambda > 1$

$$\lambda^{\lfloor (\beta+1)/2 \rfloor} + \lambda^{\lceil (\beta+1)/2 \rceil} \leqslant \lambda^{\beta+1-l} + \lambda^{l}, \qquad \text{for} \quad l = 1, \ldots, \beta \tag{3.16}$$

and hence for $l = \lfloor (\beta+1)/2 \rfloor, \lceil (\beta+1)/2 \rceil$ we have the smallest positive root of (3.15).

It remains to be seen that for this root, say λ^*, $\log \lambda^*$ is a tight bound. This, however, is a special case of the following result.

LEMMA 4. *We are given L linear equations*

(i) $b_n = g_l(b_{n-1}, \ldots, b_{n-k})$

with non-negative coefficients and the characteristic equation $\lambda^k = \psi_l(\lambda)$. *Suppose that for* $l = l^*$ *the largest root* λ^* *satisfies*

(ii) $\lambda^{*k} \leqslant \psi_l(\lambda^*)$ *for* $l = 1, \ldots, L$,

then for any positive initial values $a_0, \ldots, a_k$ *we have for the recursive equation*

(iii) $a_n = \min_{1 \leqslant l \leqslant L} g_l(a_{n-1}, \ldots, a_{n-k})$, $\lim_{n \to \infty} (1/n) \log a_n = \log \lambda^*$.

Proof. Obviously as in the previous argument $\lim_{n \to \infty} (1/n) \log a_n \leqslant \log \lambda^*$, because all coefficients are non-negative. If $\tilde{a}_t = \lambda^{*t}$, $t = 0, 1, \ldots, k$, are chosen as initial values, then the sequence $(\tilde{a}_n)_{n=0}^{\infty}$ produced by (iii) equals the sequence $(\lambda^{*n})_{n=0}^{\infty}$, because $\lambda^{*k} = \min_{1 \leqslant l \leqslant L} \psi_l(\lambda^*)$. Therefore

$$\lim_{n \to \infty} \frac{1}{n} \log \tilde{a}_n = \log \lambda^*.$$

Now for our initial values there is a $\gamma > 0$ with

$$a_t \geqslant \gamma \lambda^{*t} \qquad \text{for} \quad t = 0, \ldots, k$$

and since all coefficients are non-negative also

$$a_n \geqslant \gamma \tilde{a}_n \qquad \text{for} \quad n = 0, 1, 2, \ldots .$$

Thus

$$\lim_{n \to \infty} \frac{1}{n} \log a_n \geqslant \log \lambda^*.$$

Remark. Positivity of the initial values is essential for the result to hold, as can be seen from the

EXAMPLE. $b_n = 2b_{n-2}$, $c_n = c_{n-2} + c_{n-3}$, $a_n = \min(2a_{n-2}, a_{n-2} + a_{n-3})$. For $(a_1, a_2, a_3) = (1, 0, 1)$, $a_4 = 0$ and generally

$$a_n = \begin{cases} 0 & \text{for } n \text{ even} \\ 1 & \text{for } n \text{ odd.} \end{cases}$$

We summarize the results of Propositions 1, 2, and 3.

THEOREM 3. (i) $N^n(\infty, \beta, 0) = O^n(\infty, \beta, 0)$,

(ii) $v(\infty, \beta, 0) = \omega(\infty, \beta, 0) = \log \lambda^*$, *where* λ^* *is the largest root of* $\lambda^{\beta+1} = \lambda^{\lfloor(\beta+1)/2\rfloor} + \lambda^{\lceil(\beta+1)/2\rceil}$. *For odd* β, $\log \lambda^* = 2/(\beta+1)$.

B. *Solution for All Cases Involving* ∞*-Past*

We have determined $N^n(\infty, \beta, \varphi)$ in Theorem 2* for all $\varphi \geq \beta - 1$ and in Theorem 3 for $\varphi = 0$. We now settle the remaining cases. For this we make use of the fact

$$N^n(\infty, \beta, 0) = O^n(\infty, \beta, 0)$$

established in Theorem 3 and the idea underlying the strategy used in proving Theorem 3.

PROPOSITION 4. $N^n(\infty, \beta, \varphi) = N^n(\infty, \beta + \varphi, 0)$, *if* $\varphi \leq \beta - 1$.

Proof. Using only the first β positions in the box we see that

$$O^n(\infty, \beta + \varphi, 0) \leq N^n(\infty, \beta, \varphi),$$

and thus

$$N^n(\infty, \beta + \varphi, 0) \leq N^n(\infty, \beta, \varphi).$$

We establish the reverse inequality by showing that the strategy mentioned above can be adapted.

Let the future be $x^\varphi = (x_1, \ldots, x_\varphi)$ and let s be the usual state of the box, where $0 \leq s \leq \beta$.

Map (s, x^φ) onto $s + |x^\varphi|_1$, which can be viewed as a new state in $\{0, 1, \ldots, \beta + \varphi\}$. Since $\varphi \leq \beta - 1$ implies $\lceil(\beta + 1 + \varphi)/2\rceil \geq \lceil(2\varphi + 1)/2\rceil = \varphi + 1 > \varphi$, a suitable element, 0 or 1, to follow our optimal strategy for $N^n(\infty, \beta + \varphi, 0)$ already exists in the box.

THEOREM 3*. (i) $v(\infty, \beta, \varphi) = 1/\beta$ *for* $\varphi \geq \beta - 1$.

(ii) $v(\infty, \beta, 0) = \log \lambda^*$, *where* λ^* *is largest root of* $\lambda^{\beta+1} = \lambda^{\lceil(\beta+1)/2\rceil} + \lambda^{\lfloor(\beta+1)/2\rfloor}$.

(iii) $v(\infty, \beta, \varphi) = v(\infty, \beta + \varphi, 0)$ *for* $\varphi \leqslant \beta - 1$.

(iv) $v(\infty, \beta, \varphi) = \omega(\infty, \beta, \varphi)$ *for all* φ.

Proof. The only statement not directly contained in Theorem 2*, Theorem 3, and Proposition 4 is (iv) for $\varphi \leqslant \beta - 1$. Here the identity follows from (iii) and

$$v(\infty, \beta, \varphi) \geqslant \omega(\infty, \beta, \varphi) \geqslant \omega(\infty, \beta + \varphi, 0) = v(\infty, \beta + \varphi, 0).$$

C. *A Limit Theorem for* $\pi \to \infty$

The result below may be appreciated after one has thought about the corresponding problem for the future. There we have strong evidence for the

Conjecture. $\lim_{\varphi \to \infty} v(\pi, \beta, \varphi) \neq v(\pi, \beta, \infty)$.

THEOREM 4. *For all (finite and infinite) values of* φ

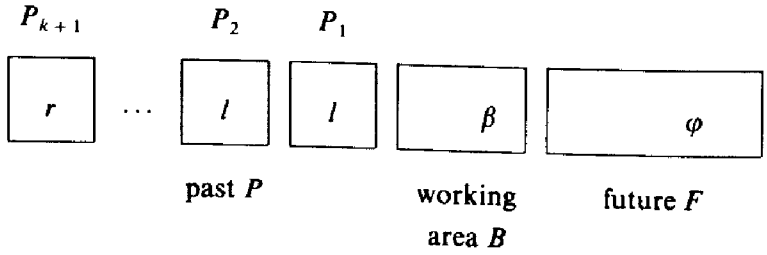
$$\lim_{\pi \to \infty} v(\pi, \beta, \varphi) = v(\infty, \beta, \varphi).$$

Proof. Since by Theorem 2*, $v(\pi, \beta, \infty) = v(\infty, \beta, \infty)$, only the cases $\varphi < \infty$ are to be treated. Here the proof is based on an interesting strategy which closely resembles our strategy in the case of knowledge of time. The finite past is used to divide the time into cycles.

We consider first the range of values

$$\begin{gathered} \varphi < \beta \qquad \pi = k\ell + 1; \\ \ell = \left\lceil \frac{\beta + 1 + \varphi}{2} \right\rceil, \ 1 \leqslant r \leqslant \ell, \ k \geqslant 1. \end{gathered} \tag{3.17}$$

It is convenient to use the following diagram, which indicates the division of time.



At the beginning either B is filled or we wait until this is the case. Since we let n tend to infinite for fixed (π, β, φ) this has no effect on the rate. We may as well assume that B is filled.

Strategy 1. The strategy is a combination of two basic procedures.

Procedure 1. Either P is not filled or it does not contain all 1's or all 0's. $\mathcal{O}$ sends ℓ 1's, if there are as many in the box and in the forseeable future, that is, in BF; otherwise he sends ℓ 0's.

(This is possible because the worst situation which could arise is that there are φ_0 0's followed by φ_1 1's in F, but then there are at least $\ell - \varphi_1 > \varphi_0$ 1's in B and after they have been sent the 1's of F have started to enter B.)

Procedure 2. P is filled with ε's (ε equals 0 or 1). $\mathcal{O}$ continues to send ε's until there are none left in B. Then he sends ℓ times the other letter.

At the beginning P is empty and Procedure 1 applies. $\mathcal{O}$ knows when it terminates, because at that time P_1 is filled for the first time with ε's only.

Then $\mathcal{O}$ applies Procedure 1 k times until $P_1, ..., P_k$ are filled. Now $\mathcal{O}$ again follows Procedure 1 until P_{k+1} is filled. Here, if not all elements in P are of the same kind, $\mathcal{O}$ completes Procedure 1 for the remaining $\ell - r$ steps; the termination occurs when every P_i has elements of one kind. Otherwise Procedure 2 applies. Clearly, from now on P is filled and one of the two procedures always applies. When they terminate every P_i has elements of one kind.

The strategy ends when n letters are sent. This generally occurs while a procedure is not completed. However, this procedure yields output sequences of the form

> every letter ε $(\varepsilon = 0, 1)$ appears in blocks of length $m \in \{\ell, 2\ell, ..., k\ell, (k+1)\ell, (k+1)\ell + 1, ..., n\}$ with the exception that the last block of a word may have any length $m \leqslant n - l$.

Let $R(n, l)$ count the number of those sequences. It satisfies a recurrence relation with the characteristic polynomial

$$\lambda^n = \lambda^{n-\ell} + \lambda^{n-2\ell} + \cdots + \lambda^{n-(k+1)\ell} + \lambda^{n-(k+1)\ell-1} + \lambda^{n-(k+1)\ell-2} + \cdots + 1. \tag{3.18}$$

This can also be written as

$$\lambda^n = \lambda^{n-\ell} + \lambda^{n-2\ell} + \cdots + \lambda^{n-(k+1)\ell}(1 + \lambda^{-1} + \lambda^{-2} + \cdots + \lambda^{-n+(k+1)\ell})$$

or equivalently as

$$\lambda^\ell = 1 + \lambda^{-\ell} + \lambda^{-2\ell} + \cdots + \lambda^{-k\ell}(1 + \lambda^{-1} + \cdots + \lambda^{-n+(k+1)\ell}). \tag{3.19}$$

Let $\rho_n(\pi)$ be its largest root and let $\rho(\pi)$ be the largest root of

$$\lambda^{\ell} = 1 + \lambda^{-\ell} + \lambda^{-2\ell} + \cdots + \lambda^{-k\ell} \cdot \frac{\lambda}{\lambda - 1}. \tag{3.20}$$

For every n the expression to the right in (3.20) exceeds that of the right-hand expression in (3.19). Therefore

$$\rho(\pi) > \rho_n(\pi) \tag{3.21}$$

and by continuity

$$\lim_{n \to \infty} \rho_n(\pi) = \rho(\pi). \tag{3.22}$$

We conclude that

$$\lim_{n \to \infty} \frac{1}{n} \log R(n, \ell) \leqslant \log \rho(\pi)$$

and, since by our strategy

$$N^n(\pi, \beta, \varphi) \leqslant R(n, \ell),$$

also

$$v(\pi, \beta, \varphi) \leqslant \log \rho(\pi). \tag{3.23}$$

Since $\pi = k\ell + r$, we derive from (3.20) that $\rho(\pi)$ is increasing in π and has a limit $\rho = \lim_{\pi \to \infty} \rho(\pi)$. It is root of

$$\lambda^{\ell} = 1 + \lambda^{-\ell} + \lambda^{-2\ell} + \cdots = \frac{1}{1 - \lambda^{-\ell}} \tag{3.24}$$

or $\lambda^{\ell} = 2$. Therefore

$$\rho = 2^{1/\ell}, \qquad \ell = \frac{\sqrt{\beta + 1 + \varphi}}{2}. \tag{3.25}$$

This and (3.23) imply

$$\lim_{\pi \to \infty} v(\pi, \beta, \varphi) \leqslant \frac{1}{\ell}. \tag{3.26}$$

Now for $\beta + \varphi$ odd we have $\ell = (\beta + 1 + \varphi)/2$.

By Proposition 4, $v(\infty, \beta, \varphi) = v(\infty, \beta + \varphi, 0)$ and by Theorem 3, $v(\infty, \beta + \varphi, 0) = 2/(\beta + 1 + \varphi)$. These facts and (3.25) give the result for $\varphi < \beta$ and $\varphi + \phi$ odd.

Case $\beta < \varphi$. Ignore the last $\varphi - \beta + 1$ positions in F. Thus we are in the case $(\beta, \varphi) = (\beta, \beta - 1)$ with $\ell = \beta$ and the previous result gives

$$\lim_{\pi \to \infty} v(\pi, \beta, \varphi) \leqslant \lim_{\pi \to \infty} v(\pi, \beta, \beta - 1) \leqslant \frac{1}{\beta},$$

which by Theorem 2* equals $v(\infty, \beta, \infty)$. Since obviously $v(\infty, \beta, \infty) \leqslant v(\pi, \beta, \beta - 1)$, the result follows here also.

Case $\varphi < \beta$, $\varphi + \beta$ even. Actually we *modify* our previous strategy so that it covers the case $\varphi + \beta$ odd as well. Define

$$\ell_0 = \left\lfloor \frac{\beta + 1 + \varphi}{2} \right\rfloor, \qquad \ell_1 = \left\lceil \frac{\beta + 1 + \varphi}{2} \right\rceil. \tag{3.27}$$

When there are ℓ_1 1's in BF, then (as before!) send them. Otherwise send ℓ_0 0's (*not*, as before, ℓ_1 0's). Repeat this until P is filled.

In the case $\varphi + \beta$ odd there is no difference from the previous strategy. But if the sum is even, then we have blocks of length ℓ_1 filled with 1's and blocks of length ℓ_0 filled with 0's. As long as both letters occur in P, $\mathcal{O}$ always knows when he has finished the task of sending such blocks. Otherwise (as before!) he continues to send the one letter occurring in P until B has none left. For the analysis of this strategy it is essential that these blocks longer than $\pi - 1$ have no effect on the asymptotics, as can be seen from (3.24), which now is to be replaced by the familiar

$$\lambda^{\beta+1+\varphi} = \lambda^{\lfloor \beta+1+\varphi)/2 \rfloor} + \lambda^{\lceil (\beta+1+\varphi)/2 \rceil}. \tag{3.28}$$

In the usual way we thus obtain

$$\lim_{\pi \to \infty} v(\pi, \beta, \varphi) = \log \lambda^*,$$

with λ^* the largest root of (6.28). By Theorem 3* therefore

$$\lim_{\pi \to \infty} v(\pi, \beta, \varphi) = v(\infty, \beta + \varphi, 0) = v(\infty, \nu, \varphi).$$

Remark. Inspection of the strategies used in Sections 5, 6, and 7 for the cases (π, β, φ) with φ equal to 0 or ∞ shows that they can all be subsumed into or be replaced by this last basic strategy. We *conjecture* this to be so for all $(\pi, \beta, 0)$ and for (π, β, φ), if π is sufficiently large compared to φ.

4. N^n *in the Knowledge of* 1*-Future or* 1*-Past*

Whereas knowledge of the ∞-future is worth more than knowledge of the ∞-past, the situation is reversed for very small values of φ and π. We

settle here the cases $(\pi, \varphi) = (0, 1)$ and $(\pi, \varphi) = (1, 0)$. Actually, $(\pi, \varphi) = (0, 1)$ is not better than $(\pi, \varphi) = (0, 0)$.

THEOREM 5. *$N^n(0, \beta, 1) = 2^n$ and thus $v(0, \beta, 1) = 1$.*

Proof. Clearly, $N^n(0, \beta, 1) \leqslant 2^n$, because $|\mathcal{Y}| = 2$. Consider now as "states" the set $\{(s, x): 0 \leqslant s \leqslant \beta; x = 0, 1\}$. A strategy f maps this set into $\{0, 1\}$, where of course $f(0, x) = 0$ and $f(\beta, x) = 1$. Let us look first at the possible transitions between states.

I. $(s, 0) \to^1 \{(s-1, 0), (s-1, 1)\}$, if $s > 0$.

II. $(s, 1) \to^1 \{(s, 0), (s, 1)\}$, if $s > 0$.

III. $(s, 0) \to^0 \{(s, 0), (s, 1)\}$, if $s < \beta$.

IV. $(s, 1) \to^0 \{(s+1, 0), (s+1, 1)\}$, if $s < \beta$.

We classify the strategies as follows:

Case 1. f follows rule I for $s > 0$ and rule IV for $s < \beta$.

Case 2. There is an s^*, $0 < s^* < \beta$,. for which f follows rule II or rule III.

Suppose first that an optimal strategy falls into the first case. Starting with a full box let $\mathcal{Y}^i(s, x)$ be the set of possible output sequences with state (s, x) after i transmissions. Then for $x = 0, 1$:

With the symbols $\dot{\cup}$ and $*$ denoting disjoint union and concatenation operations, respectively,

$$\mathcal{Y}^i(s, x) = \mathcal{Y}^{i-1}(s+1, 0) * 1 \cup \mathcal{Y}^{i-1}(s-1, 1) * 0, \qquad 0 < s < \beta \quad (4.1)$$

$$\mathcal{Y}^i(0, x) = \mathcal{Y}^{i-1}(0, 0) * 0 \cup \mathcal{Y}^{i-1}(1, 0) * 1 \quad (4.2)$$

$$\mathcal{Y}^i(\beta, x) = \mathcal{Y}^{i-1}(\beta, 1) * 1 \cup \mathcal{Y}^{i-1}(\beta - 1, 1) * 0 \quad (4.3)$$

and hence in this case

$$|\mathcal{Y}^i| = 2\,|\mathcal{Y}^{i-1}|. \quad (4.4)$$

Now note that in Case 2, when we follow, for instance, rule II for $s = s^*$, we never reach a state s' exceeding s^*; that is, we always have at least $\beta - s^*$ 0's in the box. In state $(s^*, 1)$ we send 1 according to rule II and for $s' < s^*$, as for $s = s^*$, we do not need the extra $\beta - s^*$ 0's. Therefore we can decrease the size of the box from β to s^* and follow the isomorphic strategy. By induction on the size of the box we arrive at $\beta = 1$, where we must send whatever arrives.

The situation is symmetrically the same if rule III is to be applied. Thus $N^n(0, \beta, 1) \geqslant 2^n$.

Our next result is expressed in terms of the binary entropy function h.

THEOREM 6. (a) $v(1, \beta, 0) \leqslant \sup_\delta (1-(\beta-1)\delta)\, h(\delta/(1-(\beta-1)\delta))$.

(b) $v(1, \beta, 0) \geqslant \log \psi_\beta$, *where* ψ_β *is the positive root of* $\lambda^\beta - \lambda^{\beta-1} - 1 = 0$.

(c) $\sup_\delta (1-(\beta-1)\delta)\, h(\delta/(1-(\beta-1)\delta)) = \log \psi_\beta$.

Proof. (a) The states are now $\{(x, s): x = 0, 1; 0 \leqslant s \leqslant \beta\}$. First we analyse the strategy

$$f(0, s) = 0 \qquad \text{for} \quad s < \beta$$

and

$$f(1, s) = 1 \qquad \text{for} \quad s > 0. \tag{4.5}$$

This strategy simply repeats the previous action, if this is possible. First we derive the stated upper bound, and later we show that this strategy and also this bound are optimal. Note that the sequences produced by f have the following structure:

The first letter, 0 or 1, occurs ℓ_0 times, then the other letter in $\mathscr{X}$ ℓ_1 times, then again the first letter ℓ_2 times, etc., such that for a suitable $d \leqslant \lfloor n/\beta \rfloor$ the vector $\mathbf{l} = (\ell_0, \ell_1, ..., \ell_{d+1})$ has the properties

$$\sum_{i=0}^{d+1} \ell_i = n \qquad \text{and} \qquad \ell_i \geqslant \beta \qquad \text{for} \quad i = 1, ..., d. \tag{4.6}$$

Furthermore, the cardinality $M(f)$ of the set $\mathscr{Y}(f)$ of output sequences of length n satisfies

$$M(f) \leqslant 2 \sum_{d=0}^{\lfloor n/\beta \rfloor} L(n, \beta, d), \text{ where } L(n, \beta, d) \text{ is the number of vectors } \mathbf{l} \text{ of length } d+2 \text{ satisfying (4.6).} \tag{4.7}$$

Obviously, $L(n, \beta, d)$ equals the number of vectors $(\ell_0, \ell_1 - \beta, ..., \ell_d - \beta, \ell_{d+1})$ with components in $\mathbb{N}_0$ and

$$\ell_0 + \sum_{i=1}^{d} (\ell_i - \beta) = n - \beta d. \tag{4.8}$$

This number equals $\binom{n-\beta d+d}{d}$ and therefore

$$M(f) \leqslant 2 \sum_{d=0}^{\lfloor n/\beta \rfloor} \binom{n-\beta d+d}{d}.$$

This implies

$$\nu(1, \beta, 0) \leqslant \overline{\lim_{n \to \infty}} \frac{1}{n} \log M(f)$$

$$\leqslant \max_{0 \leqslant \delta \leqslant 1/\beta} (1-(\beta-1)\delta)\, h\left(\frac{\delta}{1-(\beta-1)\delta}\right).$$

(b) First we derive the desired lower bound for $|\mathscr{Y}^n(f)|$ and then we show that f is better than all its competitors.

Let $M^n(m)$, $m = 1, ..., \beta$, be the number of output sequences of length n for which the set of possible states is either $\{0, 1, ..., m\}$ or $\{\beta - m, ..., \beta\}$. Note that these are the only sets of states occurring when we start with state set S and any memory 0 or 1. Therefore

$$M^n(f) = \sum_{m=1}^{\beta} M^n(m). \tag{4.9}$$

One readily verifies the relations

$$M^n(m) = M^{n-1}(m-1) \qquad \text{for} \quad m = 2, ..., \beta - 1 \tag{4.10}$$

$$M^n(1) = M^{n-1}(\beta) \tag{4.11}$$

$$M^n(\beta) = M^{n-1}(\beta) + M^{n-1}(\beta - 1). \tag{4.12}$$

By (4.12) and (4.10)

$$M^n(\beta) = M^{n-1}(\beta) + M^{n-1-(\beta-2)}(1)$$

and hence by (4.11)

$$M^n(\beta) = M^{n-1}(\beta) + M^{n-\beta}(\beta). \tag{4.13}$$

With $M^n(\beta) = \lambda^n$ therefore

$$\lambda^n = \lambda^{n-1} + \lambda^{n-\beta} \qquad \text{or} \qquad \lambda^\beta - \lambda^{\beta-1} - 1 = 0,$$

which implies the desired lower bound for $M^n(f)$, because by (4.9) and (4.11)

$$M^n(f) = \sum_{m=2}^{\beta-1} M^n(m) + M^n(\beta) + M^{n-1}(\beta)$$

and by (4.10) and (4.11)

$$M^n(f) = \sum_{\ell=n-\beta-1}^{n} M_\ell(\beta). \tag{4.14}$$

Now for any competitor g different from f, we distinguish between two cases.

Case 1. $\exists s^*, 0 < s^* < \beta$, $g(1, s^*) = g(0, s^*) = \varepsilon$. If $\varepsilon = 0$, then from $\{s^*, ..., \beta\}$ we never leave this set, and if $\varepsilon = 1$ this is the case for $\{0, ..., s^*\}$. Since $\max(s^*, \beta - s^*) < \beta$ the suboptimality of g follows by induction in β, the case $\beta = 1$ being trivial.

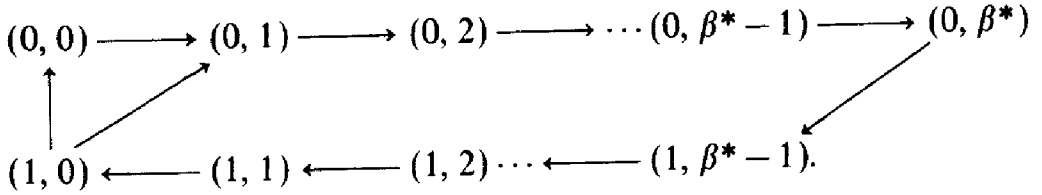
Case 2. $\exists \beta^*, 0 < \beta^* < \beta$, $g(0, \beta^*) = 1$, znf $g(1, \beta^*) = 0$. s^* is called a reversely ordered state of g. Clearly, a function G has no reversely ordered states if

$$G(1, s) \geqslant G(0, s) \qquad \text{for} \quad s = 0, 1, ..., \beta. \tag{4.15}$$

f has no reversely ordered states. By symmetry we can assume that $\beta^* + 1 \leqslant \beta - \beta^* + 1$, or that

$$1 \leqslant \beta^* \leqslant \frac{\beta}{2}. \tag{4.16}$$

We can decrease the output space only by assuming that initially we start with $(0, \{0, 1, ..., \beta^*\})$ or $(1, \{0, 1, ..., \beta^* - 1\})$. If we replace β by β^*, then strategy g behaves almost like strategy f. The difference is that there is now a transition from $(1, \beta)$ to $(0, \beta^*)$ and no transition from $(1, \beta^*)$ to $(0, \beta^* - 1)$. We just skip $(1, \beta^*)$ altogether and again only decrease the output space. We then have the following cycle, where arrows indicate transitions and we avoid drawing the loops.



To make the situation symmetric we also skip (0, 0). The possible state sets are now

$$\{1, 2, ..., m\}, \{\beta^* - m, \beta^* - 1\} \qquad \text{for} \quad m = 1, ..., \beta^*$$

and we let $M^{*n}(m)$ denote these numbers after time n. Now note that the relations (4.10) to (4.12) hold for these new starred quantities.

The initial conditions are now different, but this has no effect on the rate of growth. Therefore

$$\lim_{n \to \infty} \frac{1}{n} M^n(g) \geqslant \log \psi_{\beta^*} > \log \psi_\beta. \tag{4.17}$$

Thus, we have actually also proved that f is the only optimal strategy.

(c) The function $F(\delta)=(1-(\beta-1)\delta)\,h(\delta/(1-(\beta-1)\delta))$ is defined for $\delta\in[0,1/\beta]$. Since $F(0)=F(1/\beta)=0$ and $F(\delta)>0$ otherwise, an extremal value must be a maximum, if it is unique.

Using the definition of the binary entropy function h, we readily verify that

$$F(\delta)=-\delta\log\delta+(1-(\beta-1)\delta)\log(1-(\beta-1)\delta)$$
$$-(1-\beta\delta)\log(1-\beta\delta). \tag{4.18}$$

Since for any differentiable function G, $(G\log G)'=G'(\log G+1)$, we obtain

$$F'(\delta)=-\log\delta-(\beta-1)\log(1-(\beta-1)\delta)+\beta\log(1-\beta\delta), \tag{4.19}$$

which has a unique root δ^* satisfying

$$\delta^*(1-(\beta-1)\delta^*)^{\beta-1}=(1-\beta\delta^*)^{\beta},\qquad \delta^*\in[0,1/\beta]. \tag{4.20}$$

Since F can be written in the form

$$F(\delta)=\log\frac{(1-(\beta-1)\delta)^{1-(\beta-1)\delta}}{\delta^{\delta}(1-\beta\delta)^{1-\beta\delta}},$$

we derive with (4.20)

$$F(\delta^*)=\log\frac{1-(\beta-1)\delta^*}{1-\beta\delta^*}. \tag{4.21}$$

It remains to be seen that $(1-(\beta-1)\delta^*)/(1-\beta\delta^*)$ is a root of $\lambda^{\beta}-\lambda^{\beta-1}-1=0$. Again using (4.20), we obtain

$$\frac{(1-(\beta-1)\delta^*)^{\beta}}{(1-\beta\delta^*)^{\beta}}-\frac{(1-(\beta-1)\delta^*)^{\beta-1}}{(1-\beta\delta^*)^{\beta-1}}-1$$
$$=\frac{1-(\beta-1)\delta^*}{\delta^*}-\frac{1-\beta\delta^*}{\delta^*}-1=0.$$

Remarks. (1) We add an interesting example. Whereas by Theorem 5, $v(0,\beta,1)=1$, we have $v(0,\beta,2)<1$. To see this we consider the strategy f defined by the following diagram, in which the row-index gives the number of 1's in the future.

	0	1	⋯	$\beta-4$	$\beta-3$	$\beta-2$	$\beta-1$	β
0	0	·		0	0	0	1	1
1	0	·		0	0	1	0	1
2	0	·		0	1	0	1	1

It can be shown that

$$|\mathscr{Y}^n(f)| \leqslant |\mathscr{Y}^{n-1}(f)| + |\mathscr{Y}^{n-2}(f)| + |\mathscr{Y}^{n-2}(f)|$$

and the largest positive root λ_0 of $\lambda^3 - \lambda^2 - \lambda - 1 = 0$ is smaller than 2.

(2) We have actually proved directly that $v(1, \beta, 0) = \log \psi_\beta$. We have included (a) and (c) to see that there is an alternative expression for $v(1, \beta, 0)$ and an alternative way to derive it.

5. Results for $\alpha > 2$

A. *On* $N_\alpha^n(0, \beta, 0)$ *as* $\beta \to \infty$

The function $N_\alpha^n(0, \beta, 0)$ obeys complicated recurrence relations. We present here a result for $\alpha = 3$ and $\beta \to \infty$ which says that asymptotically in β the cases $\alpha = 2$ (see Theorem 1) and $\alpha = 3$ show the same behaviour.

THEOREM 7. $\lim_{\beta \to \infty} v_3(0, \beta, 0) = 1$.

Proof. We have the set of states

$$\mathscr{S} = \left\{(s(1), s(2), s(3)) : s(i) \in \mathbb{N}_0, \sum_{i=1}^{3} s(i) = \beta\right\}.$$

Any strategy $f: \mathscr{S} \to \{1, 2, 3\}$ can be described by a chart of triangular structure. We give an example for $\beta = 3$ in Fig. 1. If, for instance, in

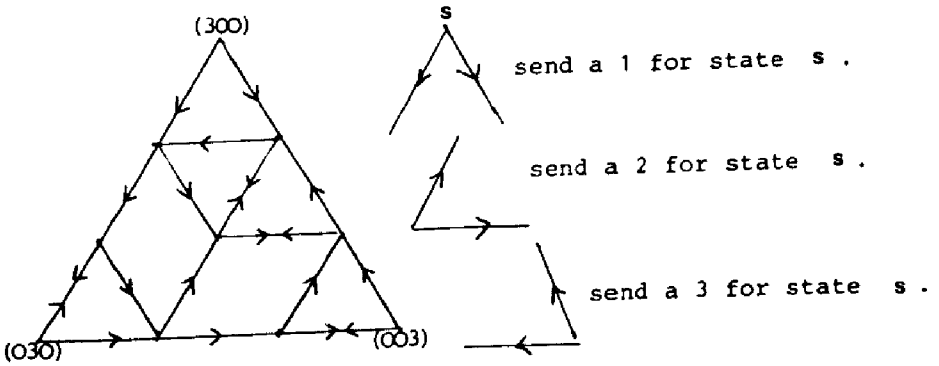


FIGURE 1

state $\mathbf{s}=(s(1), s(2), s(3))$ we send a 1, then we must consider transitions to the three states $(s(1), s(2), s(3))$, $(s(1)-1, s(2)+1, s(3))$, and $(s(1)-1, s(2), s(3)+1))$, because any one of the letters 1, 2, 3 can enter the box. Thus the arrows describe the possible transitions of states. Since every state can return to itself we have omitted the loops indicating these transitions.

Our first observation is that for any chart representing a strategy there is at least one line with arrows in opposite directions. To see this, let us start with state $(\beta, 0, 0)$. Necessarily $f(\beta, 0, 0)=1$ and thus there is a transition to the state $(\beta-1, 1, 0)$. Now either $f(\beta-1, 1, 0)=2$ and our claim is established or $f(\beta-1, 1, 0)=1$ and there is a transition to $(\beta-2, 2, 0)$. Since $f(0, \beta, 0)=2$, for some γ necessarily $f(\beta-\gamma, \gamma, 0)=1$ and $f(\beta-\gamma-1, \gamma+1, 0)=2$. Thus in state $(\beta-\gamma, \gamma, 0)$ any input word without a 3 as a letter is reproduced by f and hence

$$|\mathscr{Y}^n(f)| \geq 2^n, \qquad v_3(0, \beta, 0) \geq 1. \tag{5.1}$$

We now show asymptotic achievability of this bound by a strategy f, which corresponds to the chart shown in Fig. 2. We give the formal description:

Since $v_3(0, \beta, 0)$ is monotonically increasing in β, it suffices to consider cases where $\beta \neq 0 \bmod 3$ and $\beta > 3$. It is convenient to use the abbrevation $\gamma = \lceil \beta/3 \rceil$.

Under our assumptions we can partition the set of interior points

$$I = \{\mathbf{s}: s(i) \geq 1 \text{ for } i = 1, 2, 3\}$$

into the sets

$$I_i = I \cap \{\mathbf{s}: s(i) < \gamma, s(j) \geq \gamma \text{ with } j \equiv (i+1) \bmod 3\}$$

$(i = 1, 2, 3)$ and the sets of boundary points into the sets

$$B_i = \{\mathbf{s}: 1 \leq s(i) > \gamma, s(j) = 0 \text{ for } j \equiv (i+2) \bmod 3\}$$

and

$$C_i = \{\mathbf{s}: \gamma \leq s(i) \leq \beta, s(j) = 0 \text{ for } j \equiv (i+2) \bmod 3\}$$

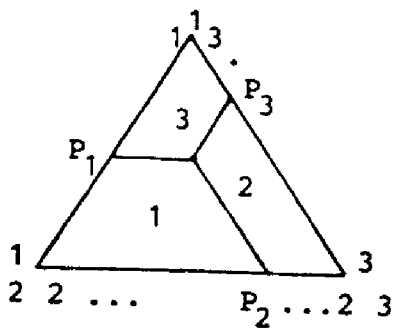


FIGURE 2

$(i=1, 2, 3)$. Set $A_i = I_i \cup B_i$. f takes the value i exactly on $D_i = A_i \cup C_i$ $(i=1, 2, 3)$. The points $P_1 = (\gamma, \beta-\gamma, 0)$, $P_2 = (0, \gamma, \beta-\gamma)$, and $P_3 = (\beta-\gamma, 0, \gamma)$ play a special role. We can enter A_1 from A_3 only via P_1 (and similarly A_2 from A_1 only via P_2, and A_3 from A_2 only via P_3).

Furthermore, starting in A_3 we can come to D_2 only via D_1 and this takes at least γ steps. Moreover, if we start in C_3, then again it takes γ steps to come to D_1. The other transitions obey analogous rules. Therefore $\mathscr{Y}^n(f)$ has the following structure:

There are three types of sequences depending on the initial state. If this state is in $A_1 \cup B_1 \cup C_2$, then we have at least γ letters from $\{1, 2\}$, at least γ letters from $\{2, 3\}$, at least γ letters from $\{1, 3\}$, etc. Therefore

$$|\mathscr{Y}^n(f)| \leqslant 3\,|L(n, \gamma)|\,2^n, \tag{5.2}$$

where $L(n, \gamma)$ is the set of sequences of numbers $(\ell_1, \ell_2, ..., \ell_{d+1})$ with $\ell_i \geqslant \gamma$ for $i=1, ..., d$ and $\sum_{i=1}^{d-1} \ell_n = n$. As in Section 4 we have the bound

$$\lim_{n\to\infty} \frac{1}{n} \log |L(n, \gamma)| \leqslant \max_{0 \leqslant \delta \leqslant 1/\gamma} (1-(\gamma-1)\delta)\, h\left(\frac{\delta}{1-(\gamma-1)\delta}\right)$$

and the right side tends to 0 as β and therefore γ tends to ∞. This and (5.2) imply the result.

B. *A Formula for* $v_\alpha(\infty, 2, 0)$

The set of all possible states is now

$$\mathscr{X}^{(2)} = \{11, 12, ..., 1\alpha; 22, ..., 2\alpha; ..., \alpha\alpha\}.$$

Let $S(y^i)$ be the set of states in $\mathscr{X}^{(2)}$ under strategy f, if the past is y^i. After y_i has left the box there is a set $A(y^i) \subset \mathscr{X}$ of possible elements in the box. Since then all letters in $\mathscr{X}$ can enter the box, $S(y^i)$ is of the form

$$S(y^i) = \bigcup_{\sigma \in A(y^i)} \{\sigma 1, ..., \sigma\alpha\}, \qquad \text{where} \quad \sigma\sigma' = \sigma'\sigma. \tag{5.3}$$

The possible sets of states in one more step under f are

$$\mathscr{S}(y^i) = \{S(y^i y) : y \in \mathscr{X}\}, \tag{5.4}$$

which depends on $A(y^i)$ and up to a permutation of $\{1, 2, ..., \alpha\}$ only on $|A(y^i)|$. For instance, if $|A(y^i)| = 1$ and w.l.o.g. $A(y^i) = \{1\}$, then $S(y^i) = \{11, 12, ..., 1\alpha\}$. If f prescribes sending 1 for all states in $S(y^i)$, then we get

$$\mathscr{S}(y^i) = \{S(y^i 1)\} = \{11, 12, ..., 1\alpha; 22, ..., 2\alpha, ..., \alpha\alpha\} = \mathscr{X}^{(2)}.$$

We denote this by $1 \to \alpha$. If $|A(y^i)| = 2$, that is, w.l.o.g. $A(y^i) = \{1, 2\}$, then $S(y^i) = \{11, 12, ..., 1\alpha; 22, ..., 2\alpha\}$. Sending 1 whenever there is a 1 in the state and otherwise 2, we obtain

$$\mathscr{S}(y^i) = \{S(y^i1), S(y^i2)\} = \{\{11, 12, ..., 1\alpha; 22, ..., 2\alpha; ... \alpha\alpha\},$$
$$\{12, 13, ..., 1\alpha; 22, ..., 2\alpha; ... \alpha\alpha\}\}.$$

We denote this by $2 \to \{\alpha, \alpha - 1\}$.

Analogously, we define

$$k \to \{\alpha, \alpha - 1, ..., \alpha - k + 1\} \qquad \text{for} \quad k = 1, 2, ..., \alpha. \tag{5.5}$$

This describes a strategy F, which we shall prove to be optimal. In order to compare it with an arbitrary strategy f, we introduce the following notion.

If $|A(y^i)| = k$, then we set

$$\beta_k^{(j)} = |\{s \in S(y_i): f(y^i, s) = j\}|; \qquad 1 \leqslant j \leqslant \alpha. \tag{5.6}$$

Thus we can assign to f and the given y^i an operation

$$k \to \{\beta_k^{(1)}, ..., \beta_k^{(\alpha)}\}. \tag{5.7}$$

It counts how often f outputs $y_{i+1} = j$ $(1 \leqslant j \leqslant \alpha)$ as s varies over $S(y^i)$.

We now prove optimality of F by induction in $m = n - i$. Because the case $m = 0$ is vacuously true, we assume optimality of F for $m < \ell$ and show optimality for $m = \ell$. If we use f at time i, then by the induction hypothesis we can use F at time $i + 1$. This results in the operations $k \to \{\beta_k^{(1)}, ..., \beta_k^{(k')}\}$ and $\beta_k^{(j)} \to \{\alpha, \alpha - 1, ..., \alpha - \beta_k^{(j)} + 1\}$ for $j = 1, ..., k'$ corresponding to $\mathscr{S}(y^i)$. Clearly, $k \leqslant k' \leqslant \alpha$.

Similarly, if we use the strategy F *twice*, then $k \to \{\alpha, ..., \alpha - k + 1\}$ and $\alpha - j + 1 \to \{\alpha, \alpha - 1, ..., j\}$ for $j = 1, ..., k$. We show that the first strategy fF can only be worse than FF and thus complete the proof.

A comparison can be made by comparing the systems of state sets after fF and FF have been performed. We obtain the trees shown in Fig. 3, where the labelling of the nodes incidates the "sizes" of the state sets and edges are labelled by the letters sent. Since by definition $\sum_{j=1}^{k'} \beta_k^{(j)} = \alpha + (\alpha - 1) + \cdots + (\alpha - k + 1)$, the number of states in "$k \times \alpha$," both trees have the same number of terminal nodes. However, for the operation of the system in the future the sizes of the state sets at the terminal nodes are relevant. Fortunately, for *any* γ there are at least as many sets of sizes greater than γ in the tree fF as in the tree FF. This fact follows from the lemma below, and since the number of successors of a node increases with increasing size of the state set, our proof is complete.

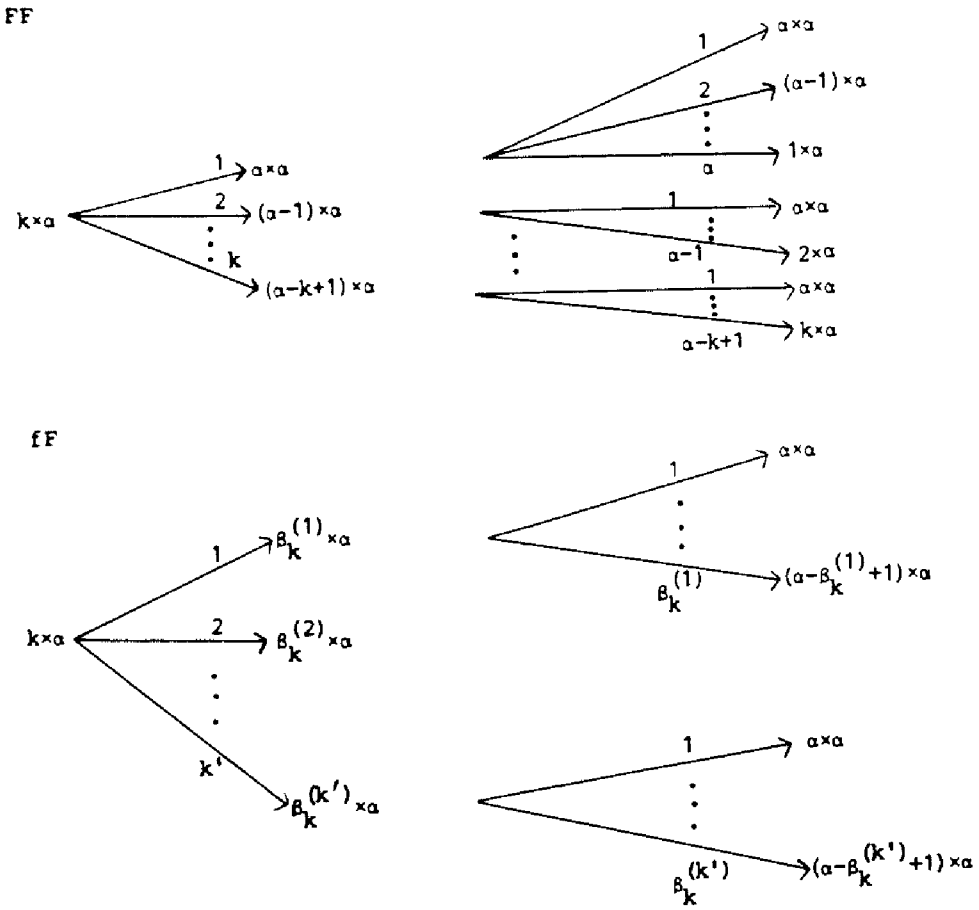


FIGURE 3

We formulate the auxiliary result in terms of matrices.

For $\varepsilon = 1, 2$ let M_ε be an $(\alpha \times k_\varepsilon)$-matrix with 0, 1 as entries and the properties

(a) $M_\varepsilon(i, j) \geqslant M_\varepsilon(i', j)$ for $i \leqslant i'$ and all j.

(b) $\sum_i M_\varepsilon(i, j) \geqslant \sum_i M(i, j')$ for $j \geqslant j'$.

(c) $\sum_{i,j} M_\varepsilon(i, j) = \alpha + (\alpha - 1) + \cdots + (\alpha - k_1 + 1)$.

Moreover,

(d) $\sum_i M_1(i, j) = \alpha - j + 1$ for $j = 1, \ldots, k_1$.

LEMMA 5. *If for any number ℓ and any ℓ columns $j_1, \ldots, j_\ell$*

(e) $\sum_{r=1}^{\ell} \sum_i M_2(i, j_r) \leqslant \alpha + (\alpha - 1) + \cdots + (\alpha - \ell + 1)$

then for every δ

$$\sum_{j=1}^{k_1} \sum_{i \leqslant \delta} M_1(i, j) \leqslant \sum_{j=1}^{k_2} \sum_{i \leqslant \delta} M_2(i, j).$$

Proof. If M_2 has α 1's in the first column, then omission of this column in both matrices reduces the problem to matrices M_ε' with $\alpha' = \alpha - 1$, $k_\varepsilon' = k_\varepsilon - 1$. In the case $k_1 = 1$ the result obviously holds.

If M_2 has fewer than α 1's in the first column, we change M_2 to M_2^* with one more 1 in the first column and one less 1 in the last column, which has the same number of 1's as the second column.

Since (a), (b), (c), and in particular (e) again hold for M_2^*, and since $\sum_{j=1}^{k_2} \sum_{j \le \delta} M_2(i, j) \ge \sum_{j=1}^{k_2} \sum_{i \le \delta} M_2^*(i, j)$, the result follows by iteration of the two reductions.

In the application the number of 1's in the ith row of M_2 equals the number of $(\alpha - i + 1) \times \alpha$ state sets.

Analysis of strategy F. Let a_i^t count the number of i's in the tth component of the output sequences. Initially, at $t = 0$, we make the convention $a_1 = 0, ..., a_{\alpha-1} = 0, a_\alpha = 1$. For $t = 1$ we have $a_1^1 = a_2^1 = \cdots = a_\alpha^1$, which can be written as

$$(a_1^1, a_2^1, ..., a_\alpha^1) = (a_\alpha, a_\alpha + a_{\alpha-1}, ..., a_\alpha + a_{\alpha-1} + \cdots + a_1) = (0, 0, ..., 1) D_\alpha,$$

where

$$D_\alpha = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \cdots & & \vdots \\ 1 & \cdots & & 1 \end{pmatrix}.$$

As can be seen from the definition of F in (5.5) (or from the diagram FF), in general

$$(a_1^{t+1}, ..., a_\alpha^{t+1}) = (a_1^t, ..., a_\alpha^t) D_\alpha$$

and thus

$$(a_1^n, ..., a_\alpha^n) = (0, 0, ..., 1)(D_\alpha)^n.$$

We have $|\mathscr{Y}^n(F)| = \sum_{i=1}^{\alpha} a_i^n$ and therefore

$$\lim_{n \to \infty} \frac{1}{n} \log |\mathscr{Y}^n(F)| = \log \psi_\alpha,$$

where ψ_α is the largest eigenvalue of D_α.

We summarise our findings.

THEOREM 8. *Strategy F is optimal.* $v_\alpha(\infty, 2, 0) = \log \psi_\alpha$, *where* ψ_α *is largest eigenvalue of* D_α.

Further, observe that for $\alpha = 2$, $\psi_2 = (1+\sqrt{5})/2$ and that by Theorem 6, $v_2(1, 2, 0) = \log((1+\sqrt{5})/2)$.

COROLLARY. *$v_2(\pi, 2, 0) = \log((1+\sqrt{5})/2)$ for all $\pi \geq 1$.*

6. On Time and Order

Recall that we know from Theorem 2* in Section 5 that

$$T^n(\pi, \beta, \varphi) = N^n(\infty, \beta, \varphi) \qquad \text{for} \quad \varphi \geq \beta - 1 \text{ and all } \pi > 0. \tag{6.1}$$

Analysis of our strategies achieving $v(\infty, \beta, \varphi)$ for $\varphi \leq \beta - 1$ (Section 3) shows that they do not make full use of the available knowledge. From knowledge of the ∞-past they use only

(a) knowledge of time,

(b) knowledge of the 1-past.

The following result is a consequence of this observation, of (6.1), and the obvious inequality $T^n(\pi, \beta, \varphi) \geq N^n(\infty, \beta, \varphi)$.

THEOREM 9. *For all $\pi \geq 1$, β, and φ*

$$\tau(\pi, \beta, \varphi) = v(\infty, \beta, \varphi).$$

Thus for $\alpha = 2$ only $\tau(0, \beta, \varphi)$ remains to be investigated. We study here the case $\varphi = 0$.

The analysis of $T^n(0, \beta, 0)$ requires a new setting of ideas. Thus far we have found only a lower bound, which is tight for $\beta = 2$. Actually, the case $\beta = 2$ can be settled much more quickly via Theorem 8, as we explain below. We included the following approach because it contains a new idea, which may be useful otherwise or may be improvable.

Since $\mathcal{O}$ knows the time, a strategy is now a sequence $f = (f_1, ..., f_n)$ of maps f_t: $\{0, 1, ..., \beta\} \to \{0, 1\}$ with $f_t(0) = 0$ and $f_t(\beta) = 1$.

Suppose that at time t the set of all possible outputs is $\{y^t(1), ..., y^t(k)\}$ and that for $j = 1, 2, ..., k$, $\mathcal{S}(y^t(j))$ is the set of all possible states for output $y^t(j)$. Then

$$|\mathcal{Y}^n(f)| = |\mathcal{Y}^{n-1}(\mathcal{S}(y^t(j)); 1 \leq j \leq k; f); \tag{6.2}$$

that is, $|\mathcal{Y}^n(f)|$ depends only on the k state sets and not on the outputs at time t.

We investigate this situation by considering more abstractly any k sets of states $U_1, ..., U_k \subset \mathcal{S} = \{0, 1, ..., \beta\}$; that is, we study $\mathcal{Y}^m(U_1, ..., U_k; f)$ with

$m=n-t$. Since the strategy is fixed we drop the letter f in our formulas. We associate with the sets $U_1,...,U_k$ sets $V_1,...,V_k$, where

$$V_\ell = \bigcup_{i_1 \neq i_2 \neq \cdots \neq i_\ell} (U_{i_1} \cap U_{i_2} \cap \cdots \cap U_{i_\ell}). \tag{6.3}$$

The lower bound mentioned is a consequence of the following inequality, which holds for all β.

LEMMA 6. $|\mathscr{Y}^m(U_1,...,U_k)| \geq |\mathscr{Y}^m(V_1,...,V_k)|$.

Proof. We proceed by induction in m.

For $m=0$ nothing is to be shown. For fixed m set $\mathscr{S}^\varepsilon = \{s \in \mathscr{S}: f_{n-m}(s)=\varepsilon\}$ and define for any $Z \subset \mathscr{S}$

$$Z^\varepsilon = Z \cap \mathscr{S}^\varepsilon; \qquad \varepsilon = 0, 1. \tag{6.4}$$

Furthermore, set

$$Z^{0*} = Z^0 \cup (Z^0+1), \qquad Z^{1*} = Z^1 \cup (Z^1-1). \tag{6.5}$$

Obviously we have

$$|\mathscr{Y}^m(U_1,...,U_k)| = |\mathscr{Y}^{m-1}(U_1^{0*},...,U_k^{0*})| + |\mathscr{Y}^{m-1}(U_1^{1*},...,U_k^{1*})| \tag{6.6}$$

and by the induction hypothesis for $m-1$

$$|\mathscr{Y}^m(U_1,...,U_k)| \geq \sum_{\varepsilon=0}^{1} |\mathscr{Y}^{m-1}(V_\ell(U_1^{\varepsilon*},...,U_n^{\varepsilon*}); 1 \leq \ell \leq k)|. \tag{6.7}$$

Since by (6.4) and (6.5)

$$U_i^{\varepsilon*} = (U_i \cap \mathscr{S}^\varepsilon) \cup ((U_i \cap \mathscr{S}^\varepsilon)+(-1)^\varepsilon),$$

we have

$$\begin{aligned} &V_\ell(U_1^{\varepsilon*},...,U_k^{\varepsilon*}) \\ &\quad = \bigcup_{i_1 \neq i_2 \neq \cdots \neq i_\ell} \bigcap_{j=1,...,\ell} ((U_{i_j} \cap \mathscr{S}^\varepsilon) \cup ((U_{i_j} \cap \mathscr{S}^\varepsilon)+(-1)^\varepsilon)) \\ &\quad \supset \left(\bigcup_{i_1 \neq i_2 \neq \cdots \neq i_\ell} \bigcap_{j=1,...,\ell} (U_{i_j} \cap \mathscr{S}^\varepsilon) \right) \\ &\quad + \{0, (-1)^\varepsilon\} = V_\ell(U_1,...,U_k)^{\varepsilon*}, \end{aligned}$$

again by (6.4) and (6.5). Now $\mathscr{Y}^{m-1}$ can decrease only if the sets of states decrease and thus by (6.7)

$$|\mathscr{Y}^m(U_1, ..., U_k)| \geqslant \sum_{\varepsilon=0}^{1} |\mathscr{Y}^{m-1}(V_\ell(U_1, ..., U_k)^{\varepsilon *}; 1 \leqslant \ell \leqslant k\}|. \quad (6.8)$$

The expression on the right side equals $|\mathscr{Y}^m(V_1, ..., V_k)|$ and the inequality is established.

THEOREM 10. *Let* $\{a_n(\beta)\}_{n=1}^{\infty}$ *be a sequence of positive integers satisfying the recurrence relation*

$$a_n(\beta) = a_{n-1}(\beta) + a_{n-\beta}(\beta)$$

and the initial conditions

$$a_n(\beta) = n+1 \qquad \text{for} \quad n = 1, 2, ..., \beta;$$

then

(i) $Tn(0, \beta, 0) \geqslant a_n(\beta)$

(ii) $T^n(0, 2, 0) = a_n(2)$

(iii) $\tau(0, 2, 0) = \log((\sqrt{5}+1)/2)$.

Proof. (i) Since for any strategy f we have $f(0) = 0$ and $f(1) = 1$, we have $\mathscr{S}^0, \mathscr{S}^1 \neq \varnothing$. One also readily verifies that

$$|\mathscr{S}^{0*} \cap \mathscr{S}^{1*}| \geqslant 2. \quad (6.9)$$

Moreover, for any non-empty $\mathscr{X} \subsetneq \mathscr{S}$

$$|\mathscr{X}^{0*} \cup \mathscr{X}^{1*}| \geqslant |\mathscr{X}| + 1. \quad (6.10)$$

Now from $|\mathscr{Y}^n(\mathscr{S})| = |\mathscr{Y}^{n-1}(\mathscr{S})^{0*})| + |\mathscr{Y}^{n-1}(\mathscr{S}^{1*})| = |\mathscr{Y}^{n-1}(\mathscr{S}^{0*}, \mathscr{S}^{1*})|$ and Lemma 6 in conjunction with the fact $\mathscr{S} = \mathscr{S}^{0*} \cup \mathscr{S}^{1*}$ we conclude that

$$|\mathscr{Y}^n(\mathscr{S})| \geqslant |\mathscr{Y}^{n-1}(\mathscr{S}, \mathscr{S}^{0*} \cap \mathscr{S}^{1*})| = |\mathscr{Y}^{n-1}(\mathscr{S})| + |\mathscr{Y}^{n-1}(\mathscr{S}^{0*} \cap \mathscr{S}^{1*})|. \quad (6.11)$$

Furthermore, again by the lemma and the monotonicity of $|\mathscr{Y}^{n-1}|$ in the state sets

$$|\mathscr{Y}^{n-1}(Z)| \geqslant |\mathscr{Y}^{n-2}(Z^{0*} \cup Z^{1*}, \varnothing)| = |\mathscr{Y}^{n-2}(Z^{0*} \cup Z^{1*})|.$$

Reiterating this argument we derive with (6.9) and (6.10)

$$|\mathscr{Y}^{n-1}(\mathscr{S}^{0*} \cap \mathscr{S}^{1*})| \geq |\mathscr{Y}^{n-\beta}(\mathscr{S})| \tag{6.12}$$

and thus from (6.11)

$$|\mathscr{Y}^{n}(\mathscr{S})| \geq |\mathscr{Y}^{n-1}(\mathscr{S})| + |\mathscr{Y}^{n-\beta}(\mathscr{S})|.$$

Verification of the initial conditions is left as an exercise.

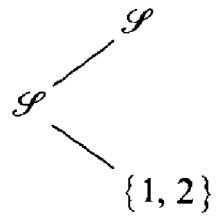
(ii) Our strategy for achieving the lower bound uses only knowledge of the parity of the time. It is described in the following diagram. The entries are the letters to be sent.

t mod 2 \ state	0	1	2
0	0	1	1
1	0	0	1

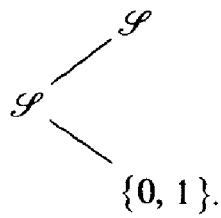
We now analyse the operation of this strategy. Starting with the set of states $\mathscr{S} = \{0, 1, 2\}$ at time $t = 1$ we send 0 for the states $s = 0$, 1 and 1 for the state $s = 2$. Then we have

$$\mathscr{S}^{0*} = \mathscr{S} \qquad \text{and} \qquad \mathscr{S}^{1*} = \{1, 2\}.$$

The transitions

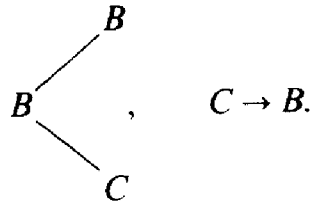


apply for all odd t. Similarily, for all even t we have the transitions



Furthermore, for even t (resp. odd t) we have the transitions $\{1, 2\} \to \mathscr{S}$ (resp. $\{0, 1\} \to \mathscr{S}$). Therefore only $\mathscr{S}$, $\{1, 2\}$, and $\{0, 1\}$ occur as state sets. The occurrence of $\{1, 2\}$ and $\{0, 1\}$ is alternating in time. They always give rise to exactly one output. Therefore, the total number of state sets after n letters have been sent equals a_n. Since $\{0, 1\}$ and $\{1, 2\}$ have isomorphic

transition rules we can denote them both by the symbol C. Let B stand for $\mathscr{S}$. Thus, the transitions of state sets can be symbolized by



$$C \to B.$$

Let b_n count the B's and c_n the C's after n letters are sent. Then we have

$$b_{n+1} = b_n + c_n, \qquad c_{n+1} = b_n, \qquad a_{n+1} = b_{n+1} + c_{n+1}. \tag{6.13}$$

Therefore $b_{n+1} = b_n + b_{n-1}$, $c_{n+1} = c_n + c_{n-1}$ and also $a_{n+1} = a_n + a_{n-1}$. All numbers are Fibonacci numbers. The initial conditions are $b_1 = c_1 = 1$, $c_2 = 2$, $c_2 = 1$, $a_1 = 2$, $a_2 = 3$.

(iii) It is well known and easy to show that, independently of the initial conditions, the rate of exponential growth of Fibonacci numbers is $\log((\sqrt{5}+1)/2)$.

Remark. The inequality $\tau_2(0, 2, 0) \geqslant \log((\sqrt{5}+1)/2)$ follows from Theorem 8, because $v_2(\infty, 2, 0) \leqslant \tau_2(0, 2, 0)$ and $(\sqrt{5}+1)/2$ is a root of

$$\det\left(\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\lambda\right) = -\lambda + \lambda^2 - 1 = 0.$$

We now collect the results on $O^n(\pi, \beta, \varphi)$, which are explicitly or implicitly contained in earlier results.

THEOREM 11. (i) $\omega(\infty, \beta, \varphi) = v(\infty, \beta, \varphi)$ *for all* φ.

(ii) $\omega(\pi, \beta, \infty) = v(\pi, \beta, \infty)$ *for all* π.

Proof. First observe that (2.7) remains true, if order is present. Therefore the proof of Theorem 2 applies literally and gives $\omega(\infty, \beta, \infty) = v(\infty, \beta, \infty)$. Since by Theorem 2*, $v(\pi, \beta, \infty) = v(\infty, \beta, \infty)$ we also have $\omega(\pi, \beta, \infty) = v(\pi, \beta, \infty)$; that is, (ii) is proved. Since again by Theorem 2*, $v(\infty, \beta, \varphi) = v(\infty, \beta, \varphi)$ for $\varphi \geqslant \beta - 1$, (i) is proved also for $\varphi \geqslant \beta - 1$.

For $\varphi \leqslant \beta - 1$ we have from Theorem 3, $\omega(\infty, \beta, 0) = v(\infty, \beta, 0)$. We also know that

$$v(\infty, \beta, \varphi) = v(\infty, \beta + \varphi, 0) \qquad \text{for} \quad \varphi \leqslant \beta - 1.$$

Since obviously $\omega(\infty, \beta, \varphi) \geqslant \omega(\infty, \beta + \varphi, 0)$, we conclude that

$$\omega(\infty, \beta, \varphi) \geqslant v(\infty, \beta, \varphi).$$

The reverse inequality is obvious.

7. Survey of Our Results for $v(\pi, \beta, \varphi)$

π, φ	$v(\pi, \beta, \varphi)$	Theorem
0, 0	1	1
0, 1	1	5
1, 0	$\sup_{\delta}(1-(\beta-1)\delta)\, h\left(\frac{\delta}{1-(\beta-1)\delta}\right)$	6
π, ∞	$1/\beta$	2*
$\infty, \leqslant \beta-1$	$\log \lambda^*$, λ^* root of $\lambda^{\beta+1+\varphi} = \lambda^{\lceil(\beta+1+\varphi)/2\rceil} + \lambda^{\lfloor(\beta+1+\varphi)/2\rfloor}$	3
$\infty, \geqslant \beta-1$	$1/\beta$	3

Finally, we emphasize that all our results are for a model in which initially all states are possible. The set of output sequences $\mathcal{Y}^n(f)$ under strategy f is the set of sequences which can be obtained as $\bigcup_s \mathcal{Y}^n(f, s)$, where s is an initial state.

Alternatively one can consider a model in which initially the box is empty. It is then filled by an arbitrary state and the output process starts. At time $n-\beta$ of the output process no new letters enter the box. The last β steps are used to clean the box. One readily verifies that this model and the model considered in this paper lead to the same rates. Mathematically the alternative model is less smooth. We report only some typical results. The quantity corresponding to N^n is marked with an asterisk.

(a) $$N^{*n}(0, \beta, 0) = 2^{n-\beta+1} + \beta - 1 \quad \text{for} \quad n \geqslant \beta.$$

(b) $$N^{*n}(\pi, \beta, \infty) = \sum_{k=0}^{n} \binom{\lfloor (n-(k \bmod \beta))/\beta \rfloor}{\lfloor k/\beta \rfloor} \quad \text{for} \quad n \geqslant 1.$$

(c) $N^{*n}(\infty, \beta, 0) = (n+1)\,|\Delta| - \sum_{\delta \in \Delta} |\delta|$, where Δ is defined as follows: let $\ell_0 = \lfloor (\beta+1)/2 \rfloor$, $\ell_1 = \lceil (\beta+1)/2 \rceil$, and let $\delta = (\ell^{(1)}, \ldots, \ell^{(*)})$, where $\ell^{(i)} = \ell_0$ or ℓ_1, and $|\delta| = \sum_{i=1}^{*} \ell^{(i)}$, then $\Delta = \{\delta : n-\beta \leqslant |\delta| \leqslant n-\beta+\ell^{(*)}\}$.

The methods of proof are the same as those in the other model. We have

$$v^*(\infty, \beta, 0) = v_2(\infty, \beta, 0).$$

Incidentally, here the tree covering problem (Proposition 3) has an analogue for trees with certain weights on the leaves, which allows an exact analysis. It is not only asymptotically but also strictly optimal always to choose ℓ_0 and ℓ_1 as lengths of two outgoing edges, as specified above.

8. Conjectures

Before we state five challenging conjectures let us first express our belief that any progress towards a determination of $N_\alpha^n(\pi, \beta, \varphi)$ for $\alpha > 2$ will depend crucially on the solution for the two subcases

$$(\pi, \beta, \varphi) = (\infty, \beta, \infty) \qquad \text{(solved only for } \alpha = 2)$$

and

$$(\pi, \beta, \varphi) = (\infty, \beta, 0) \qquad \text{(solved only for } \alpha = 2 \text{ or } \beta = 2).$$

We wonder whether general strategies in the spirit of the remark at the end of Section 3 can be found. We turn now to explicit statements.

Conjecture 1. $\lim_{\varphi \to \infty} v_2(\pi, \beta, \varphi) \neq v_2(\pi, \beta, \infty)$.

Conjecture 2. $\lim_{\beta \to \infty} v_\alpha(0, \beta, 0) = \log_2 \lceil (\alpha+1)/2 \rceil$.

(The cases $\alpha = 2$ and $\alpha = 3$ were established by Theorems 1 and 7.)

Conjecture 3. We believe that the lower bound on $T_2^n(0, \beta, 0)$ given by Theorem 10 is not tight for $\beta > 2$ and that the following strategy is optimal.

s \ $t \bmod \beta$	0	1	2	·	·	·			$\beta-1$
0	0	0	0	·	·	·			0
1	1	0	0	0	·	·	·		0
2	1	1	0	0	0	·	·	·	0
3	1	1	1	0	0	0	·	·	0
·									
·									
β	1	1	1	·	·	·	·	·	1

Conjecture 4. We view $\mathcal{X} = \{1, 2, ..., \alpha\}$ as a directed cycle and claim that an optimal strategy for $N_\alpha^n(1, \beta, 0)$ is to send the next available cyclic successor of the letter which was sent before. A letter is next successor of itself.

Conjecture 5. $\omega_2(0, \beta, 0) = v_2(1, \beta - 1, 0)$.

Actually, the inequality $\omega_2(0, \beta, 0) \leqslant v_2(1, \beta - 1, 0)$ has been established by analyzing the following strategy: For $s \in \{0, 1\}^\beta$

$$f(0, ..., 0, 1, ... 1, 0) = 0 \qquad \text{(last out)}$$

$$f(0, ..., 0, 1, ..., 1) = 0 \qquad \text{(first out)}$$

$$f(1, ..., 1, 0, ... 0, 1) = 1 \qquad \text{(last out)}$$

$$f(1, ..., 1, 0, 0, ... 0) = 1 \qquad \text{(first out)}$$

and in the other cases

$$f(s) = f(s_1, \ldots, s_\beta) = \bar{s}_1 \qquad (\text{ith out, if } i = \min\{j: s_j = \bar{s}_1\}.$$

It can be shown that

$$\omega_2(0, \beta, 0) \leqslant \log \psi_{\beta-1} = \nu_2(1, \beta - 1, 0) \qquad (\text{by Theorem 6}).$$

9. Towards a Theory of Creating Order

Contents

I. Directions of developments of our basic model for sequences.
II. Examples.
III. Ordering and source coding.
IV. Ordering, sorting, and Maxwell's demon.
V. A calculus of machines.
VI. Why do we want to create order?

I. *Directions of Developments of Our Basic Model for Sequences*

We show now that our basic model is just a prototype in a rich class of models involving rearrangements of sequences. Some lead to fascinating mathematical problems and some may be termed "semi-realistic" but still are to be expected to add to our understanding of ordering. Instead of lengthy definitions sketches of the models are given. In some cases they allow several specifications.

a. *Multiple In- and Outputs*

s in- and outputs. Instead of one object leaving and one object entering at any time instant, there may be s objects leaving and entering the box.

Varying number of outputs. Here s is again the number of objects entering the box. The number of objects leaving the box can be chosen by the organizer subject to the constraint that there be enough space in the box for the next s objects to enter.

Merging. There are numerous problems. We mention one which we find particularly neat. Suppose that there are two input sequences, both with letters from $\mathcal{X}$. $\mathcal{O}$ can look φ, say $\varphi = 1$, steps into the future in both sequences and he can choose to serve any one of the sequences, that is, let its next letter into the box. The other sequence must wait. What is the optimal rate for the output sequence?

Splitting. Again in the simplest case there is one incoming sequence, but now $\mathcal{O}$ produces two output sequences. He can at any time extend any one of those sequences according to the state of the box. What are the extremal rate pairs and the minimum of the sum of rates?

Correlation. Suppose that in the probabilistic model i.i.d. RVs X_t, $t = 1, 2, \ldots$, are of the form $X_t = (Y_t, Z_t)$, where both Y_t and Z_t take values in the same set. If $Y_t = y_t$ and $Z_t = z_t$, then at time t both y_t and z_t enter the box. $\mathcal{O}$ can output two letters and produce *one* sequence. What is the minimal mean output entropy?

b. *Objects with Special Features*

Varying-length objects. Here $\mathcal{X}$ consists of intervals of different lengths and also the working area is an interval in which intervals can be stored without overlaps. Here it is to be guaranteed that the longest interval will find space when its entrance is due.

Death, birth. Suppose that $\mathcal{X}$ consists of different animals. During the ordering process some animals die and others are born with certain probabilities. There is room for several models and questions. Similar problems arise if radioactive material is to be put in a depot. Generally, one may aim for a theory in which objects follow probabilistic transition rules anywhere in the process of ordering.

Idle objects. Suppose that one of the objects, say α, in $\mathcal{X}$ is idle. The receiver is not interested in this idle object. The organizer is free to output or ignore idle objects. On a management line "idle" stands for empty space. A different and original coding problem involving idle letters has been introduced by Roskind and Humblet (1980).

Box with exclusion rule. The previous model can be generalized as follows. Only a subset $\mathcal{S}'$ of the set of states is permitted. At any entrance time it must be guaranteed that any object entering will again lead to a permissible state. This model applies to cases where the objects consist of chemicals certain combinations of which are explosive and should therefore be avoided.

c. *Compound Objects*

Box with reaction rules. The previous model suggests another one. Suppose that certain combinations of chemicals can enter a reaction or certain pieces of garbage can be bundled, but that then this compound object can be thrown out by the organizer. Further specification of the model must state which compounds are permitted or what percentages of certain compounds are permitted.

Representatives. For any $i \in \mathcal{X} = \{1, 2, ..., \alpha\}$ there is a set of representatives $\mathcal{R}(i) \subset \{1, 2, ..., \gamma\}$. For any state $(s(1), ..., s(\alpha))$, $\sum_{i=1}^{\alpha} s(i) = \beta$, the organizer can output an $r \in \mathcal{R}(i)$ instead of an i with $s(i) \geq 1$.

Objects with many properties. Let $\mathcal{X}$ be a set of L-dimensional vectors. The following refinements of the basic model can be studied. There are L receivers and receiver ℓ distinguishes vectors only with respect to their ℓth component; that is, only this property (such as weight or color) of the objects matters to him. Accordingly he distinguishes output sequences. Thus for every strategy f each receiver has his own set of possible output sequences. What can be said about the extremal L-tuples of the cardinalities of such sets?

Exchanging parts of objects. In the previous model the components of vectors may stand for mechanical parts. $\mathcal{C}$ is now allowed to exchange parts in the box. This leads to a formidable ordering problem if one receiver is interested in all the output vectors.

d. *Errors*

Probabilistic. If $\mathcal{O}$ wants to output i and the state is s, then $w(j \mid i, s)$ is the probability that he actually outputs j. It is assumed that $w(j \mid i, s) = 0$, if $s(j) = 0$. In the probabilistic model entropy again serves as a performance criterion and in the non-probabilistic model a canonical criterion is expected cardinality.

Confusion rule. $C(i, s)$ is the set of objects which can be thrown out by $\mathcal{O}$, if he intends to send i.

Frequency rule. If i is intended, $\mathcal{O}$ acts wrongly at most λn times in time n.

Receiver can distinguish only certain objects. Let $(\mathcal{X}, \mathcal{E})$ be a graph. We say that x, x' are indistinguishable for the receiver, if $(x, x') \in \mathcal{E}$. The graph contains all loops. Let $(\mathcal{X}^n, \mathcal{E}^n)$ be the product graph, that is,

$$(x^n, x'^n) \in \mathcal{E}^n \Leftrightarrow (x_t, x'_t) \in \mathcal{E} \qquad \text{for} \quad t = 1, ..., n.$$

If now $f(\mathcal{X}^n)$ is the image of $\mathcal{X}^n$ under strategy f, then the receiver is interested in $I(f(\mathcal{X}^n))$, the minimal cardinality of a maximal independent set in $f(\mathcal{X}^n)$. The task is to determine $\min_f I(f(\mathcal{X}^n))$ for classes of strategies as defined previously.

Our probabilistic models use the classical concept of probability. Events obeying quantum probabilistic laws may be included. It is needless to dwell upon various combinations of models. We surely have missed some basic questions, but we are satisfied, if we have spread some seeds.

II. *Examples*

It is very reasonable to assume that the organizer has some memory of past actions, especially if he is a human being. Knowledge about the future seems at first glance less reasonable, especially if the input process follows no laws, as in our non-probabilistic model, or is memoryless, as in our probabilistic model. Still, there are many cases where knowledge about the future can be assumed. One may just think about a production line transporting items to a working area. A worker (organizer) can see what is on the line a certain distance ahead.

Production of goods. In many production processes several different goods are produced in succession. They can be locally ordered in a working area.

Arrival of goods and documents. A scientist receives reprints devoted to several subjects in which he is interested. He does not always take the time to put them in files according to his principles of classification. Instead he makes some local rearrangements—for instance, on his desk— before he does the final classification. Every administrator faces similar problems in dealing with documents and letters. Every salesman knows that the organization and bookkeeping of incoming goods is a formidable task.

Garbage collection. Every household produces garbage, which daily is organized such that, for instance, all paper and bottles are collected separately.

In all these examples the order creation can be viewed as a preliminary activity, which may be followed by various goal seeking actions. In the case of the management line the "organization" of parts makes it easier to put them together. The "organization" of reprints simplifies the search for a final classification. The "organization" of garbage helps in the final separation into various categories.

III. *Ordering and Source Coding*

To obtain some more specific ideas about possible concepts we look now at ordering by contrasting it to or relating it with other concepts in the familiar source coding theory of information theory, which was founded by Shannon (1949) and deals in its simplest setting with the following problem:

Having modelled the source by a sequence of independent identically distributed random variables $(X_t)_{t=1}^{\infty}$ with values in a finite set $\mathcal{X}$, we ask "How much storage space is necessary to store reliably, for fixed n, the outputs $X^n = (X_1, ..., X_n)$?" Formally, this problem can be stated as follows.

Let $\mathbb{N}$ be the set of natural numbers. For a function $f: \mathcal{X}^n \to \mathbb{N}$ we denote by $\|f\|$ the cardinality of its range. f is said to be an ε-reliable description of X^n, if there exists a function $g: \mathbb{N} \to \mathcal{X}^n$ such that

$$\Pr(g(f(X^n)) = X^n) \geq 1 - \varepsilon.$$

For any $\varepsilon \in (0, 1)$ (typically very small) the quantity

$$\log_2 N(n, \varepsilon) = \log_2 \min_f \{\|f\| : f \text{ is an } \varepsilon\text{-reliable description of } X^n\}$$

measures the number of positions needed to store X^n ε-reliably in a binary alphabet.

Shannon's souce coding theorem says that

$$\lim_{n \to \infty} \frac{1}{n} \log N(n, \varepsilon) = H(X),$$

where $H(X)$ is the entropy, that is,

$$H(X) = -\sum_{x \in \mathcal{X}} \Pr(X = x) \log \Pr(X = x).$$

Here it is important that we do not record the outputs $X^n = X_1 \cdots X_n$ themselves, but a "description" or symbol $f(X^n)$ representing them. We emphasize that, in ordering, all objects produced by the "source" are themselves to be stored. This is a crucial difference between source coding and ordering, where the material is always preserved and compression can be achieved solely by reordering matter.

However, if in addition one is interested in descriptions (records) of the storage of the objects, then this leads to a theory combining the theory of ordering and the theory of source coding. That means in particular that we aim for an ordering of objects which allows a simple description. One may think of reprints which are first ordered and then their positions are described in a file.

Suppose now that we allow descriptions which may be wrong in λn positions of our output sequence y^n. As is well known from rate-distortion theory this will shorten the descriptions significantly. Moreover, in order to obtain essentially optimal descriptions one should not encode optimal orderings but should design *a combined ordering–description procedure.* We hope for a general theory which includes both a theory of ordering and rate-distortion theory.

IV. *Ordering, Sorting, and Maxwell's Demon*

Generally, ordering is any activity which reduces alternatives or reduces entropy, if this concept is defined, that is, we are working in a probabilistic model. The sets of objects need not be sequence spaces. There is an apparent connection to sorting. In Knuth's (1973) work sorting is the task of putting a specified family $F=(a_i)_{i\in I}$ of members from a linearly ordered set $\mathscr{L}$ into its linear order. Here $a_i=a_{i'}$ may occur. Many algorithms have been designed to meet this goal under various constraints, for instance, those imposed by the machines used. Several performance criteria for storage space and running time are in use. Roughly speaking the main difference between sorting problems and the more general ordering problems we have in mind is that, in sorting, the number of alternatives is reduced to one (the linear list) whereas in ordering the number of alternatives is reduced optimally (resp. the entropy is minimized) under specified limitations. The step from sorting to the more general ordering is analogous to the generalization of ordinary channel codes (with one option for decoding) to list codes (with a list of options for decoding). Ahlswede (1973) explained that "information" can be understood as "list size reduction." "Gain of order" is an analogue to "gain of information."

The importance of ordering problems in science can perhaps best be demonstrated by their connection with the second law of thermodynamics. At what price can an intelligent being or a machine reduce entropy of an ensemble?

Generations of physicists have persuaded us that there is no way to operate on a thermodynamical system in equilibrium so that we finally receive more energy than we have put into the system. It seems to us that the only justification for this belief is that nobody was able to achieve the opposite. We have not found a convincing theoretical argument in the extensive literature!

The most famous experiment of thought in this respect is known under the name "Maxwell's demon." Brillouin (1957, p. 162) writes:

> The sorting (!) demon was born in 1871 and first appeared in Maxwell's Theory of Heat (p. 328) as "a being whose faculties are so sharpened that he can follow every molecule in his course, and would be able to do what is at present impossible to us. Let us suppose that a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being who can see the individual molecules opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A. He will, thus, whithout expenditure of work raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.

In connection with Maxwell's demon one may look at other dynamical systems, in particular those which have been invented to better understand

Boltzmann's H-theorem. The Ehrenfest urn model is the most famous system of this kind.

We think that a systematic study of ordering problems other than the Maxwell demon problem should also shed some light on the former. It seems to us that the physicists' belief in the second law, which is hallowed by failures in designing a perpetuum mobile of the second kind for more than a century, has had the negative effect that ordering problems in other models were not studied at all or at least not with enough effort. Exceptions are studies in biology. For instance, Eigen and Winkler (1975) investigate how certain life games bring about order, for instance, the ability of reproduction in certain dissipative systems, and Prigogine (1979, p. 97) mentions that cells in nervous systems perform complex operations which are based on principles similar to those of a management line.

V. *A Calculus of Machines*

The simple permuting machines which we have used can be characterized by a quadruple $\mathcal{M} = (\alpha, \pi, \beta, \varphi)$. Interesting questions arise if we study relations among several machines.

Comparisons of machines. There are some natural relations with respect to performance. We say that

$\mathcal{M} = (\alpha, \pi, \beta, \varphi)$ is *better* than $\mathcal{M}' = (\alpha, \pi', \beta', \varphi')$, if $N_\alpha^n(\pi, \beta, \varphi) \leqslant N_\alpha^n(\pi', \beta', \varphi')$ for all n.

$\mathcal{M}$ is *asymptotically better* than $\mathcal{M}'$, if $v_\alpha(\pi, \beta, \varphi) \leqslant v_\alpha(\pi', \beta', \varphi')$.

$\mathcal{M}$ is *uniformly better* than $\mathcal{M}'$, if for all n, $\mathcal{A} \subset \mathcal{X}^n$ and all strategies f' for $\mathcal{M}'$ there is a strategy f for $\mathcal{M}$ such that $|f(\mathcal{A})| \leqslant |f'(\mathcal{A})|$.

$\mathcal{M}$ is *strictly better* than $\mathcal{M}'$ if even $f(\mathcal{A}) \subset f'(\mathcal{A})$ holds.

Notions similar to the first three can be defined for probabilistic models if entropy takes the role of cardinality. Instead of using one input distribution only, one can also make comparisons for classes of such distributions.

Can the pairs of machines satisfying such relations be characterized?

Commutativity. A product $\mathcal{M} \cdot \mathcal{M}'$ can be defined by first using $\mathcal{M}$ and then using $\mathcal{M}'$ on its outputs. For several types of machines we can prove statements of the kind

$$\mathcal{M} \cdot \mathcal{M}' \text{ is asymptotically better than } \mathcal{M}' \cdot \mathcal{M}.$$

Commutativity in this asymptotic sense holds only in exceptional cases.

Whenever we have several machines available with which to create order it is certainly important to know in which sequence we should use them.

Instead of operating in serial it may sometimes be better to use the machines in parallel. Further questions arise if the machines have different *costs* of operation.

VI. *Why Do We Want to Create Order?*

The foregoing discussion was meant to increase our awareness of the complexity and range of the topic "ordering." It may be instructive to contrast these ideas with those held by Shannon (1956) and those held by Wiener (1955) more than 30 years ago.

There is still another way to look at things. Why do we want to create order?

Answers have been given explicitly and implicitly in the Introduction and in the discussion of Maxwell's demon. We have also mentioned a short record as a possible goal. This makes it easier to instruct someone about the positions of objects in a sequence. There is a related, though different, goal: we want to make the task of searching as easy as possible. This is not achieved automatically if the organizer optimally reduces the set of alternatives. Instead, this reduction must be best suited for the available searching algorithms or machines used for the search. Here much work is to be done. The quantities defined in the Introduction are addressed to only one aspect of the ordering problem.

RECEIVED February 10, 1988; FINAL MANUSCRIPT RECEIVED June 12, 1989

REFERENCES

AHLSWEDE, R. (1973), Channel capacities for list codes, *J. Appl. Probab.* **10**, 824–836.
AHLSWEDE, R. (1979), Coloring hypergraphs: A new approach to multiuser source coding, I, *J. Combin. Inform. System Sci.* **4**, 76–115.
AHLSWEDE, R. (1980), Coloring hypergraphs: A new approach to multiuser source coding, II, *J. Combin. Inform. System Sci.* **5**, 220–268.
AHLSWEDE, R., AND KASPI, A. (1987), Optimal coding strategies for certain permuting channels, *IEEE Trans. Inform. Theory* **33**, 310–314.
AHLSWEDE, R., AND WEGENER, I. (1987), "Suchproblème," Teubner, Stuttgart; English ed. "Search Problems," Wiley, New York.
AHLSWEDE, R., AND ZHANG, Z. (1989), Contributions to a theory of ordering for sequence spaces, *Probl. Control Inform. Theory* **18** (4), 197–221.
BERGER, T. (1971), "Rate-Distortion Theory: A mathematical Basis of Data Compression," Prentice–Hall, Englewood Cliffs, NJ.
BRILLOUIN, L. (1957), "Science and Information Theory," Academic Press, San Diego, CA.
EIGEN, M., AND WINKLER, R. (1975), "Das Spiel," Piper, Münich.

KNUTH, D. E. (1973), "The Art of Computer Programming," Vol. 3, "Sorting and Searching," Addison-Wesley, Reading, MA.

KOBAYASHI, K. (1987), Combinatorial structure and capacity of the permuting relay channel, *IEEE Trans. Inform. Theory* **33**, No. 6, 813–826.

PRIGOGINE, I. (1979), "From Being to Becoming—Time and Complexity in Physical Sciences," Freeman, San Francisco.

ROSKIND, J. A. AND HUMBLET, P. A. (1980), "Protocols for Encoding Idle Characters in Data Streams," Master's thesis, MIT, June.

SHANNON, C. E. (1948), A mathematical theory of communication, *Bell. System Tech. J.* **27**, 379–424, 623–657.

SHANNON, C. E. (1956), The bandwagon, *IRE Trans. Inform. Theory* **IT-2**, No. 1, 3.

WIENER, N. (1955), What is information theory? *IRE Trans. Inform. Theory* **IT-1**, No. 3, 2.

ZIV, J. (1978), Coding theorems for individual sequences, *IEEE Trans. Inform. Theory* **IT-24**, 405–512.

Printed in Belgium