Communication Systems Which Minimize Coding Noise

By BROCKWAY McMILLAN

(Manuscript received May 28, 1969)

The problem of minimizing coding or quantizing noise in a communication system is posed in a general setting. It is shown that if the messages to be transmitted are sample sequences drawn from a discrete-time random process meeting a certain simply stated criterion of "randomness" and if there exists a quantized communication system which is optimal in that it introduces a minimum amount of coding noise, then this optimal system can be realized using a transmitter of special form. Specifically, the optimum transmitter is one which quantizes each message sample according to a scheme that depends only upon the quantized material already transmitted, rather than upon the (unquantized) material that has been previously offered for transmission. It follows that only digital storage is required at the transmitter or receiver. If the receiver is limited, a priori, *to have only a given finite amount of storage, and if the system is optimum within this constraint, the transmitter need have only the same amount of storage.*

I. INTRODUCTION: THE MODEL

Shannon's theory of communication, shows how to defeat noise introduced in a communication medium by restricting the repertoire of transmitted signals to a discrete set.[1] If the messages to be transmitted are not already in an appropriately discrete form, noise in the medium is then eliminated only at the expense of noise, here called coding noise, caused by the failure of the restricted family of available signals to represent faithfully the full family of possible messages. The amount of coding noise introduced is of course subject to control by design.

This paper considers one aspect of the problem of minimizing coding noise. Noise in the medium is not considered. The paper limits attention to systems in which the random process representing the message is a discrete-time or sampled-data process. The sampling noise caused by creating such a process out of a continuous-time process is not considered.

The problem of selecting a coding scheme that maximizes the rate of communication over a noisy channel is not considered. Rather, the paper starts at the point that a coding scheme has been found, that is optimum according a fairly general criterion of fidelity. What is then shown is that the transmitter and receiver—encoder and decoder—of the system are of a special form.

A Q-coded communication system is defined by a discrete set Q and by three jointly distributed random processes, $\{x_n, q_n, y_n \mid n = 0, \pm 1, \pm 2, \cdots\}$. For purposes of this paper, the set Q will be either

(*i*) the set $\{1, 2, \cdots, M\}$, where M is a given positive integer > 1, or
(*ii*) the set $\{1, 2, 3, \cdots\}$ of all positive integers.

The process $\{x_n\}$ represents periodic samples derived from the message offered for transmission, each x_n is a real random variable. $\{q_n\}$ represents the transmitted signals; for each n, q_n is a random variable, taking values from the set Q and measurable on the sample space of $\{x_n, x_{n-1}, x_{n-2}, \cdots\}$. That is, for each n, the value of the integer variable q_n depends only upon, and is determined (apart perhaps from events of probability zero) by the present and past of the message. $\{y_r\}$ represents the version of the message reconstructed at the receiver; for each n, y_n is a real random variable measurable on the sample space of $\{q_n, q_{n-1}, q_{n-2}, \cdots\}$. Therefore for each n, y_n depends only upon, and is determined (apart perhaps from events of probability zero) by the present and past of the transmitted signal.

The model at this point is very general. It provides that at each time, n a discrete valued random variable q_n be generated in some way out of the material $\{x_n, x_{n-1}, x_{n-2}, \cdots\}$ then available from the message process, and that subsequently at the receiver a y_n be generated out of the material $\{q_n, q_{n-1}, \cdots\}$ there currently available. If all three processes $\{x_n, q_n, y_n\}$ are stationary we can call the system stationary. The question of stationarity does not enter in what follows.

What remains to be specified in this model is that in some sense the process $\{y_n\}$ is to represent the process $\{x_n\}$. At the start it appears natural to consider three cases; it develops that two are simply special cases of the third, one of them not interesting in the framework of this paper.

We start with a given sequence $\{\psi_n \mid n = 0, \pm 1, \pm 2, \cdots\}$ of functions, in which each ψ_n is a real valued Borel measurable function $\psi_n(x, y)$ of the real variables x, y. The use of a sequence $\{\psi_n\}$ here is a largely decorative generality that costs nothing. The conventional case is that in which all ψ_n are the same function ψ. These functions define a fidelity criterion as follows:

Case (i), the *delay-free case*:

Here we choose to regard y_n as a replica of x_n, and evaluate our communication system at each time n by the quantity

$$E\{\psi_n(x_n, y_n)\}, \tag{1}$$

where E denotes expectation over the message ensemble.

Case (ii), the case of *fixed delay*:

Here we are given a fixed integer $d \geqq 0$ and we choose to regard y_n as a replica of x_{n-d}, thus allowing q_n to take advantage not only of $\{x_{n-d}, x_{n-d-1}, \cdots\}$ (the present and past of x_{n-d}) but also of $\{x_n, x_{n-1}, \cdots, x_{n-d+1}\}$ (a limited span of the "future" of x_{n-d}) in representing x_{n-d}. Here the criterion relative to x_{n-d} is (by a convention we will use with respect to indices)

$$E\{\psi_n(x_{n-d}, y_n)\}. \tag{2}$$

If $d = 0$, this case reduces to case i.

Case (iii), *block encoding* with cycle time c:

This is the situation that arises naturally in Shannon's theory. We are given a fixed integer $c \geqq 1$, and the transmission process is repetitive with a cycle of length c. By a choice of time origin, we can describe it as follows. Let Q_1 be a discrete set with $M_1 < \infty$ members. At time 0 the transmitter examines $\{x_0, x_{-1}, \cdots\}$ and generates a Q_1-discrete variable which we shall call $\hat{q}_0$. At time c, the transmitter then examines $\{x_c, x_{c-1}, \cdots\}$ and produces $\hat{q}_1$; the process repeats with period c. For transmission, the random variable $\hat{q}_0$ is encoded into the string $\{q_c, q_{c-1}, \cdots, q_1\}$ of random variables each being Q-discrete, where $M^c \geqq M_1$. At time c, all of $\hat{q}_0, \hat{q}_{-1}, \cdots$ are available at the receiver, being represented by the sequence $\{q_c, q_{c-1}, q_{c-2}, \cdots\}$. From these, the sequence $\{y_{2c-1}, y_{2c-2}, \cdots, y_c\}$ is generated, representing $x_0, x_{-1}, \cdots, x_{-c+1}$, respectively. We think of these y's as being presented to the output of the receiver in the order of their indices, y_c at time c, and so on.

If one follows through the functional dependencies here, he sees that indeed the processes $\{x_n, q_n, y_n\}$ are so related that each q_n depends at most upon $\{x_n, x_{n-1}, \cdots\}$, and each y_n at most upon $\{q_n, q_{n-1}, \cdots\}$. Indeed, except at times which recur with period c, q_n is not "up to date," depending in fact only on x's strictly prior to x_n. Similarly, y_n is only periodically up to date; at other times it depends only upon q's that are actually earlier than q_n.

In the situation as just described, the criterion of fidelity becomes $E\{\psi_n(x_{n-2c+1}, y_n)\}$. Case iii is then also a special case of case ii, in which

$d = 2c - 1 \geqq 1$. What makes it special is that in case *ii*, q_n and y_n are permitted to be up to date at each value of n, however in case *iii* the block coding process restricts the currency of the data upon which most of the q's and y's depend.

Actually, case *iii* as just described will turn out not to be covered, in general, by the theorems to be proved. This happens because, as is later be stated more precisely, we are interested only in communication systems that minimize (2) for each n, in comparison with all possible competing systems. Clearly, to impose the restrictions immanent in case *iii* upon one's reportoire of coding schemes limits the domain within which a minimum is to be sought. The system that brings about an absolute minimum is simply not, in general, to be found in this restricted domain.

The previous observation is not to be entered as a criticism of Shannon's theory. Typically, in a noisy medium, it is necessary to use a highly redundant encoding $\{q_c, q_{c-1}, \cdots, q_1\}$ to represent $\hat{q}_0$, so that the inefficiencies (as measured by expression 2) that are imposed by the block-coding process are needed in order to ensure that the y_n in (2) is an approximately error free replica of x_{n-d}. We must remember that (2) measures the noise introduced by the coding process, not by the noisy medium. It is interesting to a designer only if the latter noise has been eliminated. The price of this elimination is that one may not be able to minimize (2) in competition with systems that are not restricted to be of block coding form.

A true engineering solution to the problems reflected in the remarks immediately above would consider (2) in which the expectation is taken over the joint ensemble of message and noise. The solution should balance coding noise against channel noise at, say, a fixed delay, to minimize (2). This paper is very far from solving such a problem.

It does not follow that the results of this paper are without interest in the search for coding schemes to eliminate noise. Given a Q-coded communication system which does minimize (2), the $\{q_n\}$ process is in digital form. This $\{q_n\}$ process can then be redundantly encoded according to Shannon's theory, and recovered with few errors (and typically much delay) at the receiver. The $\{y_n\}$ process then results (perhaps delayed) and has few errors. Then (2) does measure the total amount of noise introduced in this operation.

II. STATEMENT OF RESULTS

Given the message process $\{x_n\}$, the sequence $\{\psi_n\}$, and the delay $d \geqq 0$, a Q-coded communication system $\{x_n, q_n, y_n\}$ will be called

$\{\psi_n, d\}$-optimal if

(*i*) For each $n = 0, \pm 1, \pm 2, \cdots$

$$E\{| \psi_n(x_{n-d}, y_n) |\} < \infty, \tag{3}$$

and

(*ii*) For any other Q-coded communication system $\{x_n, q_n', y_n'\}$,

$$E\{\psi_n(x_{n-d}, y_n)\} \leqq E\{\psi_n(x_{n-d}', y_n')\}, \tag{4}$$

for each $n = 0, \pm 1, \pm 2, \cdots$.

The simplest result of this paper is of such a form as to illustrate the nature of all of the results. We define a class K of functions ψ, and a class, here called *CCD*, of message processes $\{x_n\}$, such that the following theorem is true.

Theorem 1: Let $\{x_n, q_n, y_n\}$ *be a given* Q-*coded communication system that is* $\{\psi_n, 0\}$*-optimal. If each* $\psi_n \in K$, $n = 0, \pm 1, \pm 2, \cdots$, *and if* $\{x_n\} \in$ CCD, *then each* q_n *is equal with probability one to a random variable measurable on the sample space of* $\{x_n, q_{n-1}, q_{n-2}, \cdots\}$.

The force of this theorem is that it simplifies, in principle at least, the requirements for memory at the transmitter. Only the digital sequence $\{q_{n-1}, q_{n-2}, \cdots\}$ need be in storage at time n. The proof of the theorem will also develop a standard structure for the optimum transmitter difficult to summarize easily in a theorem.

The definition of the class K is long and is deferred to Section III. Suffice it here to say that K is a large class that includes the conventional

$$\psi^1(x, y) = | x - y |, \qquad \psi^2(x, y) = (x - y)^2,$$

and any other continuous strictly increasing function of ψ^1.

We define *CCD*, and a related class *CCDf*, thus:

CCD consists of those processes $\{x_n\}$ such that: for each $n = 0, \pm 1, \pm 2, \cdots$, if z is a random variable measurable on the sample space of $\{x_{n-1}, x_{n-2}, \cdots\}$, then the probability that $z = x_n$ is zero:

$$P\{z = x_n\} = 0. \tag{5}$$

CCDf consists of those processes $\{x_n\}$ such that: for each $n = 0, \pm 1, \pm 2, \cdots$, if **A** is a finite Borel field or the completion of a finite Borel field, and if z is a random variable measurable on the smallest Borel field containing **A** and the sample space of $\{x_{n-1}, x_{n-2}, \cdots\}$, then (5) holds.

Read *CCD* as "continuous conditional distribution." If $\{x_n\} \varepsilon$ *CCD* and if x_n has a conditional distribution given $\{x_{n-1}, x_{n-2}, \cdots\}$, that distribution must be continuous.

We now define a more restricted class of Q-coded communication systems and a corresponding notion of optimality.

Given an integer $m \geqq 0$, a Q-coded communication system $\{x_n, q_n, y_n\}$ will be said to have *decoder memory span m* if for each $n = 0, \pm 1, \pm 2, \cdots$ y_n is measurable on the sample space of $\{q_n, q_{n-1}, \cdots, q_{n-m}\}$.

A Q-coded communication system $\{x_n, q_n, y_n\}$ will be called $\{\psi_n, d, m\}$-optimal if it has decoder memory span m, if (3) holds for every n, and if (4) holds for every n and for every $\{x_n, q'_n, y'_n\}$ which has decoder memory span m.

In the case of $\{\psi_n, d, m\}$ optimality, then, the competition is restricted to systems with decoder memory span m. We can put $m = \infty$ to refer to the case of $\{\psi_n, d\}$ optimality defined earlier.

Perhaps our most surprising result is that case *ii* of our model, which includes case *i* as a special case, is also included in case *i*. This is shown by Theorem 2.

Theorem 2: Let $\{x_n, q_n, y_n\}$ *be a given* Q*-coded communication system that is* $\{\psi_n, d\}$*-optimal. If each* $\psi_n \varepsilon K$, $n = 0, \pm 1, \pm 2, \cdots$, *if* M, *the number of elements of* Q, *is finite, and if* $\{x_n\} \varepsilon$ *CCDf, then each* q_n *is equal with probability one to a random variable measurable on the sample space of* $\{x_{n-d}, q_{n-1}, q_{n-2}, \cdots\}$. *Furthermore, the system* $\{x_n, q'_n, y'_n\}$, *where*

$$q'_n = q_{n+d}, \qquad n = 0, \pm 1, \pm 2, \cdots, \tag{6}$$
$$y'_n = y_{n+d},$$

is a Q*-coded communication system that is* $\{\psi'_n, 0\}$ *optimal, where*

$$\psi'_n = \psi_{n+d}, \qquad n = 0, \pm 1, \pm 2, \cdots. \tag{7}$$

Finally, we state a theorem that includes the two preceding ones.

Theorem 3: Let $\{x_n, q_n, y_n\}$ *be a given* Q*-coded communication system that is* $\{\psi_n, d, m\}$*-optimal. If each* $\psi_n \varepsilon K$, $n = 0, \pm 1, \pm 2, \cdots$, *if* $M < \infty$, *and if* $\{x_n\} \varepsilon$ *CCDf, then each* q_n *is equal with probability one to a random variable measurable on the sample space of* $\{x_{n-d}, q_{n-1}, \cdots, q_{n-m}\}$ ($\{x_{n-d}\}$ *if* $m = 0$). *The system as defined by* (6) *is a* Q*-coded communication system with decoder memory span m that is* $\{\psi'_n, 0, m\}$*-optimal, where* ψ'_n *is given by* (7). *If, in the initial hypotheses,* $d = 0$, *then it suffices that* $\{x_n\} \varepsilon$ *CCD and the restriction* $M < \infty$ *may be removed. If* $m < \infty$, *the hypothesis* $\{x_r\} \varepsilon$ *CCDf may be replaced by:*

For each n $= 0, \pm 1, \pm 2, \cdots$, *if* z *is a random variable that takes only finitely many values, then* $P\{x_n = z\} = 0$.

Theorem 1 shows the basic facts about measurability in the present context. Theorem 2 adds the fact that delay $d > 0$ gains no advantage (since the "future" of x_{n-d} is not known at the receiver, even if it is at the transmitter). Finally, Theorem 3 includes these facts and shows that a limitation on the memory span of the receiver allows a corresponding simplification of the transmitter.

In the proofs of these theorems it is seen that they are true for classes of process slightly larger than *CCD* or *CCDf*. In particular, the final conclusion of Theorem 3 opens the case of finite memory span to any process $\{x_n\}$ that has a little additive nonsingular Gaussian noise in each sample.

III. THE CLASS K

The class K of cost functions allowed by these theorems can be very general. The definition below seems more inclusive than is called for by the applications I can think of; at the cost of elaboration, it can be enlarged further.

We let K be the class of all functions $\psi(x, y)$ of two real variables x, y with the following properties.

(*i*) $\psi(x, y)$ is continuous;
(*ii*) for all $x, y,$ $\psi(x, y) \geqq 0;$
(*iii*) for all $x,$ $\psi(x, x) = 0;$
(*iv*) for each y, there are at most countably many solutions x to the equation

$$\psi(x, y) = 0, \tag{8}$$

in the sense that: there exist Borel measurable functions $g_k(y)$, $k = 1, 2, 3, \cdots$, such that if (8) holds, then for some k, $x = g_k(y)$.

v) If $y_1 \neq y_2$, there are at most countably many solutions to the equation

$$\psi(x, y_1) = \psi(x, y_2), \tag{9}$$

in the sense that: there exist Borel measurable functions $f_k(y, z)$, $k = 1, 2, 3, \cdots$ such that if (9) holds and if $y_1 \neq y_2$, then for some k, $x = f_k(y_1, y_2)$.

It follows from this definition that $\psi^1 \in K$, where $\psi^1(x, y) = | x - y |$. Then also $\psi^2 \in K$, where $\psi^2(x, y) = (x - y)^2$. Similarly any other continuous strictly monotone function of ψ^1 is also in K. In all of these instances, (8) has the unique solution $y = x$, and (9) has a unique solution given by $2x = y_1 + y_2$.

IV. PROOFS

Let $\{\Omega, \mathbf{B}, P\}$ be a probability space: A set Ω of points ω, a Borel field $\mathbf{B}$ of subsets of Ω, and a probability measure P on $\mathbf{B}$ with respect to which $\mathbf{B}$ is complete. This probability space is assumed given and fixed.

A random variable x is a real-valued function $x(\omega)$ defined on Ω and measurable $\mathbf{B}$.

If $\mathbf{F} \subseteq \mathbf{B}$ is a Borel field, a random variable x is said to be *essentially measurable* $\mathbf{F}$ if x is equal with probability one to a random variable x' which is measurable $\mathbf{F}$. If $\mathbf{F}$ is complete, such an x is then itself measurable $\mathbf{F}$.

If $\mathbf{F} \subseteq \mathbf{B}$ is a Borel field and x a random variable, $\{x\} \vee \mathbf{F}$ denotes the smallest Borel field such that: x is measurable $\{x\} \vee \mathbf{F}$ and $\mathbf{F} \subseteq \{x\} \vee \mathbf{F}$.

A random variable taking its values in the set Q will be called Q-discrete.

Denote by $[x \mid q, \mathbf{F} \mid y, \mathbf{G}]$ a mathematical object of the following kind:

x is a random variable,

q is a Q-discrete random variable,

$\mathbf{F}$ is a Borel field, $\mathbf{F} \subseteq \mathbf{B}$, and q is essentially measurable on the field determined by $\mathbf{F}$ and the sample space of x,

y is a random variable,

$\mathbf{G}$ is a Borel field, $\mathbf{G} \subseteq \{x\} \vee \mathbf{F}$, and y is essentially measurable on the field determined by $\mathbf{G}$ and the sample space of q.

For convenience let $CQAx$ ("conditionally quantized approximation to x") denote the class of all objects of the kind described, based on the given probability space $\{\Omega, \mathbf{B}, P\}$, the given x, and the given set Q.

Given a Q-coded communication system $\{x_n, q_n, y_n\}$, given a delay d and a memory span m, let $\mathbf{X}_{n,d}$ be the sample space of the selection $\{x_n, x_{n-1}, \cdots\}$ of random variables from which the specific variable x_{n-d} has been deleted. Let $\mathbf{Q}_{n,m}$ be the sample space of the random variables $\{q_{n-1}, q_{n-2}, \cdots, q_{n-m}\}$. Then it is easy to see that $\{x_n, q_n, y_n\}$ is a Q-coded communication system with decoder memory span m if and only if for each $n = 0, \pm 1, \pm 2, \cdots$

$$[x_{n-d} \mid q_n, \mathbf{X}_{n,d} \mid y_n, \mathbf{Q}_{n,m}] \in CQAx_{n-d} .$$

Given ψ, a $[x \mid q, \mathbf{F} \mid y, \mathbf{G}] \varepsilon CQAx$ will be called weakly ψ-optimal if:

(*i*) $E\{\mid \psi(x, y) \mid\} < \infty$,

(*ii*) If random variables q' and y' are such that $[x \mid q', \mathbf{F} \mid y', \mathbf{G}] \varepsilon CQAx$, then $E\{\psi(x, y)\} \leqq E\{\psi(x, y')\}$.

The qualifier "weakly" in this definition signals the fact that the fields **F** and **G** are not allowed to vary in the competition for optimality.

Lemma 1: If $\{\mathrm{x_n}, \mathrm{q_n}, \mathrm{y_n}\}$ *is a* Q*-coded communication system with decoder memory span* m, *and if* $\{\mathrm{x_n}, \mathrm{q_n}, \mathrm{y_n}\}$ *is* $\{\psi_\mathrm{n}, \mathrm{d}, \mathrm{m}\}$*-optimal, then for each* n $[\mathrm{x_{n-d}} \mid \mathrm{q_n}, \mathbf{X}_{\mathrm{n,d}} \mid \mathrm{y_n}, \mathbf{Q}_{\mathrm{n,m}}]$ *is weakly* ψ_n*-optimal.*

Proof: Fix an n; for convenience identify it as $n = 0$. Suppose that we are given random variables q' and y', which we shall here call q_0' and y_0', such that

$$[x_{-d} \mid q_0', \mathbf{X}_{0,d} \mid y_0', \mathbf{Q}_{0,m}] \varepsilon CQAx_d .$$

Define a new Q-coded communication system $\{x_n, q_n', y_n'\}$ thus:

For $n < 0$, $q_n' = q_n$, $y_n' = y_n$;

For $n = 0$, q_0' and y_0' are those above;

For $n > 0$, $q_n' = 1$ and $y_n' = 0$.

That this is a Q-coded communication system with decoder memory span m follows at once from the definitions. Furthermore, the sample space of $\{q_{-1}', q_{-2}', \cdots q_m'\}$ is $\mathbf{Q}_{0,m}$. Because $\{x_n, q_n, y_n\}$ is $\{\psi_n, d, m\}$-optimal, we conclude that $E\{\mid \psi_0(x_{-d}, y_0) \mid\} < \infty$ and that $E\{\psi_0(x_{-d}, y_0)\} \leqq E\{\psi_0(x_{-d}, y_0')\}$.
These, however, prove that $[x_{-d} \mid q_0, \mathbf{X}_{0,d} \mid y_0, \mathbf{Q}_{0,m}]$ is weakly ψ_0-optimal. Clearly this proof can be repeated for any other value of n.

The proof of this lemma indicates, deliberately, the force of the notion of $\{\psi_n, d, m\}$-optimality for $\{x_n, q_n, y_n\}$. The competing communication system $\{x_n, q_n', y_n'\}$ used in the proof sacrificed all reasonable behavior for $n > 0$, yet was still allowed to compete at $n = 0$. In particular, notice that even if $\{x_n, q_n, y_n\}$ is stationary, it must compete with nonstationary systems designed to excel at only one value of n. The theorems of Section II are not proved for stationary systems which are known only to minimize each $E\{\psi_n(x_{n-d}, y_n)\}$ against competing systems drawn from the class of stationary systems.

Given a Borel field $\mathbf{G} \subset \mathbf{B}$, we define $CCD(\mathbf{G})$ analogously to CCD: $CCD(\mathbf{G})$ is the class of all random variables x such that:

If z is a random variable measurable **G**, then $P\{x = z\} = 0$.

The results of this paper all derive from Theorem 4.

Theorem 4: Let $[x \mid q, \mathbf{F} \mid y, \mathbf{G}] \varepsilon$ CQAx *and suppose that it is weakly* ψ*-optimal. If* Q *is a finite set, or if* ψ *is Borel measurable and for each* x *is bounded from below, then there exists a* Q*-discrete random variable* q' *and a random variable* y' *such that*

(*i*) $[x \mid q', \mathbf{G} \mid y', \mathbf{G}] \varepsilon$ CQAx,
(*ii*) $\psi(x, y') = \psi(x, y)$ *with probability one. In particular, also, the object i is weakly* ψ*-optimal.*
If $\psi \varepsilon$ K *and* x ε CCD(G) *then also*
(*iii*) $q' = q$ *with probability one, and*
(*iv*) $y' = y$ *with probability one.*

It then follows that the given q *is essentially measurable on the Borel field* $\{x\}$ v**G**, *determined by* **G** *and the sample space of* x.

We wish to use the given $[x \mid q, \mathbf{F} \mid y, \mathbf{G}]$ as a model for some

$$[x_{n-d} \mid q_n , \mathbf{X}_{n,d} \mid y_n , \mathbf{Q}_{n,m}]$$

in a Q-coded communication system. Conclusions i and ii show that for any given n we can find a q'_n essentially measurable $\{x_{n-d}\} \vee \mathbf{Q}_{n,m}$ and a y'_n such that, according to the criterion defined by ψ, y'_n represents x_{n-d} as well as y_n did. Without conclusion iii, however, the substitution of q'_n for q_n can alter the subsequent Borel fields $\mathbf{Q}'_{n+k,m}$, $k \geqq 0$, to the point that we are no longer sure that $[x_{n+k-d} \mid q'_{n+k} , \mathbf{X}_{n+k,d} \mid y'_{n+k} , \mathbf{Q}'_{n+k,m}]$, $k > 0$ is weakly ψ_{n+k}-optimal. Without iii, therefore, one cannot apply Theorem 4 to prove the other theorems.

It is convenient now to invoke a lemma which is a simple theorem from measure theory. The lemma provides a standard form for the variables q and y of an object $[x \mid q, \mathbf{F} \mid y, \mathbf{G}] \varepsilon$ $CQAx$.

Theorem 2: Given a Q*-discrete random variable* q *and a Borel field* **G**, *if* y *is a random variable measurable on the Borel field determined by* **G** *and the sample space of* q, *then there exist random variables* $\{z_p , p \varepsilon Q\}$ *such that*

(*i*) *each* z_p *is measurable* **G** *and*
(*ii*) *for each* $\omega \varepsilon \Omega$, *if* $q(\omega) = p$ *then* $y(\omega) = z_p(\omega)$.

Conversely, of course, given $\{Z_p , p \varepsilon Q\}$, *each measurable* **G**, *any* **y** *defined by ii is measurable on the field determined by* **G** *and the sample space of* q.

The proof of this lemma consists in showing that the class of random variables of the type of y above, as the $\{z_p , p \varepsilon Q\}$ are selected arbitrarily from the class of variables measurable **G**, exhausts the class of all random variables measurable on the Borel field determined by **G** and the sample space of q. The proof is a straightforward exercise in measure theory and is omitted.

To begin the main argument, given $[x \mid q, \mathbf{F} \mid y, \mathbf{G}] \in CDAx$ and a Borel measurable function $\psi(x, y)$, if for each x $\psi(x, y)$ is bounded from below, or if Q is a finite set, we can define the random variable

$$\xi(\omega) = \inf_{r \in Q} \psi(x(\omega), z_r(\omega)).$$

Then ξ is measurable $\{x\} \vee \mathbf{G}$.

Given $p \in Q$ and $r \in Q$, we define sets T_p^*, T_{pr}, T_p by

$$T_p^* = \{\omega \mid \psi(x(\omega), z_p(\omega)) = \xi(\omega)\},$$

$$T_{pr} = \{\omega \mid \psi(x(\omega), z_p(\omega)) = \psi(x(\omega), z_r(\omega))\},$$

$$T_p = T_p^* - \bigcup_{\substack{r \neq p \\ r \in Q}} T_{pr} .$$

Clearly each of these sets is measurable $\{x\} \vee \mathbf{G}$. T_p^* is the set where the index p minimizes $\psi(x, z_p)$, and T_p is that subset of T_p^* where this minimizing index is unique. It follows that if $r \neq p$ then

$$T_p \wedge T_r^* = \emptyset, \tag{10}$$

and as a consequence, $T_p \wedge T_r = \emptyset$, $r \neq p$.

Clearly

$$T_{pr} = T_{rp} .$$

Also

$$T_r^* \wedge T_{pr} = T_p^* \wedge T_{pr} , \tag{11}$$

since either side is the set where an index minimizing $\psi(x, z_s)$ can be equal either to p or to r.

In terms of these sets, the argument to be used can be outlined briefly. First, one shows that the T_p^* essentially cover Ω, in the sense that there is a null set N such that

$$\Omega - N = \bigcup_{p \in Q} T_p^* . \tag{12}$$

This follows without argument, and with $N = \emptyset$, if Q is finite; it results from ψ-optimality in general.

Second, by definition

$$T_p^* - T_p \leqq \bigcup_{\substack{r \in Q \\ r \neq p}} T_{pr} . \tag{13}$$

Third, one observes that for $p, r \varepsilon Q$ and $p \neq r$, T_{pr} consists of the set S_{pr}

$$S_{pr} = \{\omega \mid z_p(\omega) = z_r(\omega)\}$$

plus a disjoint remainder $T_{pr} - S_{pr}$. The hypothesis $x \varepsilon CCD(\mathbf{G})$ allows one to show that this remainder is a null set. Over the set S_{pr}, on the other hand, the information about x conveyed by the family $\{z_p , p \varepsilon Q\}$ is redundant. The hypothesis of ψ-optimality can then be violated, unless S_{pr} is also a null set. It follows then that each T_{pr} is a null set, and from (12) and (13) then that the T_p partition Ω apart from a null set. From this the full theorem follows quickly.

To proceed with (12), given $p \varepsilon Q$, let N_p be the set

$$N_p = \{\omega \mid q(\omega) = p\} \wedge \{\Omega - \bigcup_{r \varepsilon Q} T_r^*\}.$$

Fix an $\omega \varepsilon N_p$; then $y(\omega) = z_p(\omega)$ but $\omega \not\varepsilon T_p^*$, so that $\xi(\omega) < \psi(x(\omega), z_p(\omega))$. It follows that there is some $r \varepsilon Q$, $r \neq p$, such that

$$\psi(x(\omega), z_r(\omega)) < \psi(x(\omega), z_p(\omega)), \tag{14}$$

and indeed, since Q is bounded from below, that there is a least such r, call it $r_p^*(\omega)$. Notice that N_p is measurable on the Borel field determined by the sample space of $\{x\}$, by $\mathbf{F}$, and by $\mathbf{G}$. Since $\mathbf{G} \subseteq \{x\} \vee \mathbf{F}$, it follows that N_p is measurable $\{x\} \vee \mathbf{F}$. That subset R_{pk} of N_p where $r_p^*(\omega) = k$ is empty if $k = p$; otherwise

$$R_{pk} = N_p \wedge \{\omega \mid \psi(x(\omega), z_1(\omega)) < \psi(x(\omega), z_p(\omega))\} \quad \text{if} \quad k = 1 \neq p,$$

$$R_{pk} = N_p \wedge \{\omega \mid \psi(x(\omega), z_k(\omega)) < \psi(x(\omega), z_p(\omega))\} \wedge$$

$$\cdot \bigcap_{j=1}^{k-1} \{\omega \mid \psi(x(\omega), z_j(\omega)) \geqq \psi(x(\omega), z_p(\omega))\} \quad \text{if} \quad k > 1, \quad k \neq p.$$

It follows from these equalities that R_{pk} and r_p^* are measurable $\{x\} \vee \mathbf{F}$.

We now define the Q-discrete random variable q' by

If $p \varepsilon Q$ and $\omega \varepsilon N_p$, $q'(\omega) = r_p^*(\omega)$;

If $\omega \varepsilon \Omega - \bigcup_{p \varepsilon Q} N_p$, then $q'(\omega)$ is the least value of $r \varepsilon Q$ such that $\omega \varepsilon T_r^*$.

Since the N_p cover the complement of $\bigcup_r T_r^*$, and since Q is bounded from below, this defines $q'(\omega)$ for each $\omega \varepsilon \Omega$; clearly q' is Q-discrete. Given $k \varepsilon Q$, the set where $q' \leqq k$ consists of the union of

$$\bigcup_{p \varepsilon Q} R_{pk}$$

with the set V_k, where

$$V_1 = T_1^*$$

$$V_k = (\Omega - T_1^*) \wedge \cdots \wedge (\Omega - T_{k-1}^*) \wedge T_k^*, \qquad k > 1.$$

Since each V_k is measurable $\{x\} \vee \mathbf{G} \subseteq \{x\} \vee \mathbf{F}$, it follows that q' is measurable $\{x\} \vee \mathbf{F}$. Furthermore, over $\Omega - \bigcup_{p \varepsilon Q} N_p$, q' is equal to a random variable that is measurable $\{x\} \vee \mathbf{G}$, since each V_k is measurable on this latter field.

We now define the random variable y' by

$$y'(\omega) = z_{q'(\omega)}(\omega), \qquad \omega \varepsilon \Omega.$$

Then y' is measurable on $\mathbf{G}$ and the sample space of q'. It follows that $[x \mid q', \mathbf{F} \mid y', \mathbf{G}] \varepsilon CQAx$, and from the hypothesis of weak ψ-optimality then that

$$E\{\psi(x, y)\} \leqq E\{\psi(x, y')\}. \tag{15}$$

But now we claim that for all $\omega \varepsilon \Omega$

$$\psi(x(\omega), y'(\omega)) \leqq \psi(x(\omega), y(\omega)). \tag{16}$$

First, if $\omega \varepsilon N_p$, we have

$$\begin{aligned} \psi(x(\omega), y'(\omega)) &= \psi(x(\omega), z_{r_p^*(\omega)}(\omega)) < \psi(x(\omega), z_p(\omega)) \\ &= \psi(x(\omega), y(\omega)), \end{aligned} \tag{17}$$

the inequality being by definition of r_p^*. Therefore strict inequality prevails in (16) for $\omega \varepsilon \bigcup_{p \varepsilon Q} N_p$. Consider now an $\omega \varepsilon (\Omega - \bigcup_{r \varepsilon Q} N_r) \wedge \{\omega' \mid q'(\omega') = p\}$. For this ω we have $\omega \varepsilon T_p^*$ and $\psi(x(\omega), y'(\omega)) = \psi(x(\omega), z_p(\omega)) \leqq \psi(x(\omega), z_r(\omega))$ for any $r \varepsilon Q$, by definition of T_p^*. But then (16) follows for this ω because $y(\omega) = z_r(\omega)$ for some $r \varepsilon Q$.

Now from (16), by taking expectations, we conclude the inequality opposite in sense to (15), hence (15) is an equality, and (16) is then an equality with probability one. Therefore *ii* of Theorem 4 is proved. Now by (17), (16) is a strict inequality over $N = \bigcup_{p \varepsilon Q} N_p$. Hence this latter set is a null set. Therefore *i* of Theorem 4 is proved, since q' is equal, over the complement of N, to a variable that is measurable $\{x\} \vee \mathbf{G}$, as we noted earlier. Finally, since

$$\Omega - \bigcup_{p \varepsilon Q} T_p^* = \bigcup_{r \varepsilon Q} N_r = N$$

the T_p^* essentially cover Ω. This is (12), as was to be proved.

It would be possible at this point to invoke the hypotheses $\psi \varepsilon K$

and $x \varepsilon CCD(\mathbf{G})$ to conclude *iv* of the Theorem. It will be more efficient to prove *iii* and *iv* together. To do so requires, as our earlier outline suggests, that we examine the sets $T_p^* \wedge T_{pr}$ over which redundancy prevails (because on $T_p^* \wedge T_{pr}$ either of z_p or z_r, where $r \neq p$, could be used to define the same value of y minimizing $\psi(x, y)$.

We have concluded (12), that except for $\omega \varepsilon N$, a null set, for each ω there is at least one $p \varepsilon Q$ such that $\xi(\omega) = \psi(x(\omega), z_p(\omega))$, that is, the minimizing index is uniquely p for $\omega \varepsilon T_p - N$.

Now define, as earlier, for $r \neq p$,

$$S_{pr} = \{\omega \mid z_p(\omega) = z_r(\omega)\}.$$

Then if $\omega \varepsilon T_{pr} - S_{pr}$, we have

$$\psi(x(\omega), z_p(\omega)) = \psi(x(\omega), z_r(\omega)), \qquad z_p(\omega) \neq z_r(\omega).$$

Since $\psi \varepsilon K$, it follows that for some $k = 1, 2, \cdots$ we have

$$x(\omega) = f_k(z_p(\omega), z_r(\omega)). \tag{18}$$

Now let A_{kpr} be the set of all ω such that (18) holds. We have just showed that

$$T_{pr} - S_{pr} \subseteq \bigcup_{k=1}^{\infty} A_{kpr}. \tag{19}$$

But now, since f_k is Borel measurable and each z_p is measureable $\mathbf{G}$, (18) constrains x on A_{kpr} to be equal to a random variable measurable $\mathbf{G}$. Since $x \varepsilon CCD(\mathbf{G})$, then A_{kpr} is a subset of some null set,

$$P\{A_{kpr}\} = 0, \qquad k = 1, 2, \cdots,$$

and

$$\sum_{k=1}^{\infty} P\{A_{kpr}\} = 0.$$

This last with (19) makes $P\{T_{pr} - S_{pr}\} = 0$. Indeed, finally, since Q is countable,

$$P\{\bigcup_{p \varepsilon Q} \bigcup_{\substack{r \varepsilon Q \\ r \neq p}} (T_{pr} - S_{pr})\} = 0.$$

It is important later that by definition, S_{pr} is measurable $\mathbf{G}$ and therefore that, by (19), T_{pr} is essentially measurable $\mathbf{G}$.

We now define a new Q-discrete random variable q'' and a corresponding y''. The construction depends upon an arbitrarily chosen

$p_0 \varepsilon Q$ and an arbitrarily chosen real number a, although the notation will not emphasize this dependence. Later it will be shown that $q'' = q'$ and $y'' = y'$ each with probability one, so that the dependence upon p_0 and a is not essential.

Fix a $p_0 \varepsilon Q$ and select a real number a. Define the random variable $z''_{p_0}(\omega)$ by:

$$\text{if} \quad \omega \varepsilon \bigcup_{\substack{r \varepsilon Q \\ r \neq p_0}} T_{p_0 r}\,, \qquad z''_{p_0}(\omega) = a,$$

$$\text{otherwise,} \quad z''_{p_0}(\omega) = z_{p_0}(\omega).$$

Then $z''_{p_0}(\omega)$ is measurable **G**. Define

$$z''_p = z_p\,, \qquad p \varepsilon Q, \qquad p \neq p_0\,.$$

Then certainly each z''_p, $p \varepsilon Q$, is measurable **G**. Define the Q-discrete random variable $q''(\omega)$ by

$$\text{If } \omega \varepsilon T_{p_0} \vee [(T^*_{p_0} - T_{p_0}) \wedge \{\omega' \mid \psi(x(\omega'), a) < \psi(x(\omega'), z_{p_0}(\omega'))\}]$$

then $q''(\omega) = p_0$;

$$\text{if} \quad \omega \varepsilon (T^*_{p_0} - T_{p_0}) \wedge \{\omega' \mid \psi(x(\omega'), a) \geqq \psi(x(\omega'), z_{p_0}(\omega'))\}$$

then $q''(\omega)$ is the least value of $p \varepsilon Q$ such that $p \neq p_0$ and $\omega \varepsilon T^*_p$;

$$\text{if} \quad \omega \varepsilon \Omega - T^*_{p_0}\,, \text{ then } q''(\omega) = q'(\omega).$$

It is easily seen that this defines q'' for all $\omega \varepsilon \Omega$.

We now define the random variable y'' by $y''(\omega) = z''_{q''(\omega)}(\omega)$. Then y'' is measurable on **G** and the sample space of q'', so that by construction $[x \mid q'', \mathbf{F} \mid y'', \mathbf{G}] \varepsilon CQAx$. Applying the hypothesis of weak ψ-optimality, we conclude that

$$\int_\Omega [\psi(x, y'') - \psi(x, y)]\, dP = E\{\psi(x, y'')\} - E\{\psi(x, y)\} \geqq 0. \tag{20}$$

We now partition the domain Ω of integration into the four sets

$$A_1 = T_{p_0}$$

$$A_2 = (T^*_{p_0} - T_{p_0}) \wedge \{\omega \mid \psi(x(\omega), a) < \psi(x(\omega), z_{p_0}(\omega))\},$$

$$A_3 = (T^*_{p_0} - T_{p_0}) \wedge \{\omega \mid \psi(x(\omega), a) \geqq \psi(x(\omega), z_{p_0}(\omega))\},$$

$$A_4 = \Omega - T^*_{p_0}\,.$$

That this is a partition follows from the definition and the fact, already proved, that $T_{p_0} \subseteq T^*_{p_0}$. We consider the four resulting integrals individually, in the order of the listing.

If $\omega \in T_{p_0}$ then either $\omega \in T_{p_0} \wedge N$, or $\omega \in T_{p_0} - N$. We may ignore the first case. For the second, by definition of T_{p_0}, if $r \neq p_0$

$$\psi(x(\omega), z_{p_0}(\omega)) < \psi(x(\omega), z_r(\omega)). \tag{21}$$

Also, by definition

$$\omega \neq \bigcup_{\substack{r \in Q \\ r \neq p_0}} T_{p_0 r} ,$$

and therefore by definition $z''_{p_0}(\omega) = z_{p_0}(\omega)$, and $q''(\omega) = p_0$. Then

$$\psi(x(\omega), y''(\omega)) = \psi(x(\omega), z''_{p_0}(\omega)) = \psi(x(\omega), z_{p_0}(\omega))$$

and from the inequality (21) we conclude that the integrand

$$\psi(x(\omega), y''(\omega)) - \psi(x(\omega), y(\omega)) < 0,$$

since $y(\omega)$ is equal to some $z_r(\omega)$, $r \in Q$. Hence the integral over A_1 is not positive.

If $\omega \in A_2$, then by definition $q''(\omega) = p_0$ and

$$y''(\omega) = z''_{p_0}(\omega).$$

Again, we ignore the contribution of $A_2 \wedge N$. If $\omega \in A_2 - N$ then by (13),

$$\omega \in \bigcup_{\substack{r \in Q \\ r \neq p_0}} T_{p_0 r} .$$

Then by definition $z''_{p_0}(\omega) = a$. Hence, the integrand

$$\begin{aligned}\psi(x(\omega), y''(\omega)) &- \psi(x(\omega), y(\omega)) \\ &= [\psi(x(\omega), a) - \psi(x(\omega), z_{p_0}(\omega))] + [\psi(x(\omega), z_{p_0}(\omega)) - \psi(x(\omega), y(\omega))].\end{aligned}$$

The first bracket on the right is <0 by definition of A_2, and the second is $\leqq 0$ because $\omega \in T^*_{p_0}$ and by definition of $T^*_{p_0}$ we have $\psi(x(\omega), z_{p_0}(\omega)) \leqq \psi(x(\omega), z_r(\omega))$ for all $r \in Q$; among the latter is $\psi(x(\omega), y(\omega))$. Hence the second integral is not positive, and its integrand is strictly negative.

Now consider $\omega \in A_3$. We ignore the integral over $A_3 \wedge N_1$. If $\omega \in A_3 - N_1$, then $q''(\omega) = p \neq p_0$ and $\omega \in T^*_p$ for some $p \in Q$. For this ω we have

$$\psi(x(\omega), y''(\omega)) = \psi(x(\omega), z''_p(\omega)) = \psi(x(\omega), z_p(\omega)) \leqq \psi(x(\omega), z_r(\omega))$$

for all $r \in Q$; here the first equality is by definition of y'', the second by definition of z_p'' since $p \neq p_0$, and the inequality is by definition of T_p^*. But the inequality makes the integrand in (20) $\leqq 0$, since $y(\omega) = z_r(\omega)$ for some $r \varepsilon Q$. Therefore, the integral over A_3 is not positive.

Over A_4, the integrand of (20) is

$$[\psi(x, y'') - \psi(x, y')] + [\psi(x, y') - \psi(x, y)].$$

The second bracket vanishes with probability one by *ii* of Theorem 4, already proved. The first bracket is

$$\psi(x(\omega), z''_{q'(\omega)}(\omega)) - \psi(z(\omega), z_{q'(\omega)}(\omega))$$

and this vanishes for all $\omega \varepsilon A_4$ by the definitions because over A_4, $\xi(\omega) < \psi(x(\omega), z_{p_0}(\omega))$ so that $q'(\omega) \neq p_0$; therefore by definition $z''_{q'(\omega)}(\omega) = z_{q'(\omega)}(\omega)$.

We conclude from these calculations that the integral (20) cannot be positive. By (20), therefore, the integral vanishes. But the argument showed that the integrand was $\leqq 0$ with probability one, hence indeed, the integrand vanishes with probability one:

$$\psi(x, y'') = \psi(x, y) \quad \text{with probability one.}$$

In particular, over A_2, the integrand was strictly < 0. Therefore A_2 has probability zero. We shall now exploit this fact.

In the argument above, a was any real number. Let $\{a_n\}$ be a countable dense set of real numbers and let

$$W_n = \{\omega \mid \psi(x(\omega), a_n) < \psi(x(\omega), z_{p_0}(\omega))\}.$$

We have just proved that $P\{A_2\} = 0$, which is to say that we could have proved, for each n, that

$$P\{(T_{p_0}^* - T_{p_0}) \wedge W_n\} = 0.$$

Then also

$$N_2 = \bigcup_n (T_{p_0}^* - T_{p_0}) \wedge W_n$$

is a null set. Now if $\omega \varepsilon N_2$, then $\omega \varepsilon T_{p_0}^* - T_{p_0}$ and also there is some number a_n such that

$$\psi(x(\omega), a_n) < \psi(x(\omega), z_{p_0}(\omega)). \tag{22}$$

Conversely, if $\omega \varepsilon T_{p_0}^* - T_{p_0}$ and there is a number a_n such that (22) is true, then $\omega \varepsilon N_2$. Therefore if $\omega \varepsilon (T_{p_0}^* - T_{p_0}) - N_2$, then for every number a_n we have

$$\psi(x(\omega), a_n) \geqq \psi(x(\omega), z_{p_0}(\omega)). \tag{23}$$

Given an $\omega \varepsilon (T^*_{p_0} - T_{p_0}) - N_2$, choose a sequence $a_n \rightarrow x(\omega)$. Assume that $\psi \varepsilon K$. Then ψ is continuous and from (23) we have

$$0 = \psi(x(\omega), x(\omega)) = \lim \psi(x(\omega), a_n) \geqq \psi(x(\omega), z_{p_0}(\omega)) \geqq 0.$$

Notice, incidentally, that it suffices here for each x that $\psi(x, y)$ be continuous for y in some neighborhood of x. This is an example of one way in which K can be enlarged.

From this and item *iv* in the definition of K, there is some integer K such that

$$x(\omega) = g_k(z_{p_0}(\omega)). \tag{24}$$

Let C_k be the set of all ω such that (24) holds. Since g_k is Borel measurable, over C_k, (24) constrains x to be equal to a function measurable $\mathbf{G}$. If $x \varepsilon CCD(\mathbf{G})$, then C_k is a null set. But we have just showed above that

$$(T^*_{p_0} - T_{p_0}) - N_2 \subseteq \bigcup_{k=1}^{\infty} C_k .$$

Therefore

$$P\{T^*_{p_0} - T_{p_0}\} = 0.$$

Since p_0 was arbitrary, this can be proved for each $p_0 \varepsilon Q$; therefore from (12) the T_p, $p \varepsilon Q$ essentially cover Ω. We proved along with definitions that the T_p are pairwise disjoint, hence they partition $\Omega - N_3$, where N_3 is some null set.

We continue the argument using the selected p_0. For $\omega \varepsilon \Omega - N_3$, either $\omega \varepsilon T_{p_0}$ or $\omega \varepsilon T_r$ where $r \varepsilon Q$ but $r \neq p_0$. In this latter case, however, as we proved with the definitions, $\omega \varepsilon \Omega - T^*_{p_0}$; then by definition $q''(\omega) = q'(\omega)$. If $\omega \varepsilon T_{p_0}$, by the definitions $q''(\omega) = q'(\omega) = p_0$. Therefore

$$q'' = q' \quad \text{with probability one.} \tag{25}$$

Furthermore we know that if $\omega \varepsilon T_p$, then $q'(\omega) = p$. From (25)

$$y''(\omega) = z''_{q'(\omega)}(\omega). \tag{26}$$

If $\omega \varepsilon \Omega - T_{p_0}$, except at most on a null set we have $q''(\omega) \neq p_0$ and from (26) and the definition of z''_p

$$y''(\omega) = z''_{q'(\omega)}(\omega) = z_{q'(\omega)}(\omega) = y'(\omega), \qquad \omega \varepsilon (\Omega - T_{p_0}) \wedge N_5 \tag{27}$$

where N_5 is a null set. Now if $\omega \varepsilon T_{p_0} - N$, we showed earlier that $z''_{p_0}(\omega) = z_{p_0}(\omega)$. Hence the equalities in (27) hold for $\omega \varepsilon T_{p_0} - N$ as well, so that

$$y'' = y' \quad \text{with probability one.} \tag{28}$$

Equalities (25) and (28) free the constructions from any dependence, except on a null set, upon the initially selected p_0 and a. We need the Theorem to make identification with q and y.

Let S_p be that subset of T_p where $q(\omega) \neq p$. Then if $\omega \varepsilon S_p$, by definition of T_p,
$\psi(x(\omega), y'(\omega)) = \psi(x(\omega), z_p(\omega)) < \psi(x(\omega), z_{q(\omega)}(\omega)) = \psi(x(\omega), y(\omega))$.
From *ii* of Theorem 4, then, $P\{S_p\} = 0$, and $P\{\bigcup_{r \varepsilon Q} S_r\} = 0$. Since the T_p, $p \varepsilon Q$, essentially partition Ω, it follows that $q' = q$ with probability one, and at once that $y(\omega) = z_{q(\omega)}(\omega) = z_{q'(\omega)}(\omega) = y'(\omega)$ with probability one. These conclusions are *iii* and *iv* of the Theorem, the proof of which is now complete.

To prove Theorem 1, let $\{x_n, q_n, y_n\}$ be a given Q-coded communication system that is $\{\psi_n, 0\}$ optimal. Given n, by Lemma 1,

$$[x_n \mid q_n, \mathbf{X}_{n,0} \mid y_n, \mathbf{Q}_{n,\infty}] \varepsilon CQAx_n$$

and is weakly ψ_n-optimal. If $\psi_n \varepsilon K$ and $x_n \varepsilon CCD(\mathbf{Q}_{n,\infty})$, Theorem 4 proves that q_n is measurable on $\{x_n\} \vee \mathbf{Q}_{n,\infty}$. But $\mathbf{Q}_{n,\infty}$ is the sample space of $\{q_{n-1}, q_{n-2}, \cdots\}$, and is therefore contained in the sample space of $\{x_{n-1}, x_{n-2}, \cdots\}$, since by hypothesis $\{x_n, q_n, y_n\}$ is a Q-coded communication system. The hypothesis $\{x_n\} \varepsilon CCD$ of Theorem 1 then implies that for the given n, $x_n \varepsilon CCD(\mathbf{Q}_{n,\infty})$, and Theorem 4 establishes Theorem 1.

Turning to Theorem 3, let $\{x_n, q_n, y_n\}$ be a given Q-coded communication system with decoder memory span m, and suppose that it is $\{\psi_n, d, m\}$-optimal. By Lemma 1, then, given n, $[x_{n-d} \mid q_n, \mathbf{X}_{n,d} \mid y_n, \mathbf{Q}_{n,m}] \varepsilon CQAx_{n-d}$ and is weakly ψ_n-optimal. By the hypotheses of Theorem 3, $\psi_n \varepsilon K$, and $\{x_n\} \varepsilon CCDf$. Consider $\mathbf{Q}_{n,m}$, the sample space of $\{q_{n-1}, q_{n-2}, \cdots, q_{n-m}\}$. Suppose first that $m > d$; then this sample space is the smallest Borel field which contains both the sample space of $\{q_{n-1}, \cdots, q_{n-d}\}$ and that of $\{q_{n-d-1}, \cdots, q_{n-m}\}$. Since $M < \infty$, the first of these is a finite field, and the second is a subfield of $\{x_{n-d-1}, x_{n-d-2}, \cdots\}$ (since $\{x_n, q_n, y_n\}$ is indeed a Q-coded communication system). The hypothesis $\{x_n\} \varepsilon CCDf$ then implies that $x_n \varepsilon CCD(\mathbf{Q}_{n,m})$. If $m \leqq d$, the subfield of $\{x_{n-d-1}, \cdots\}$ is empty, but the reasoning and conclusion are still valid. Then Theorem 4 applies and we conclude that q_n is measurable on the sample space of $\{x_{n-d}, q_{n-1}, \cdots, q_{n-m}\}$. This is the first conclusion of Theorem 3. We note now that a weaker hypothesis than $\{x_n\} \varepsilon CCDf$ could suffice here. Indeed, if $m < \infty$, it is sufficient that: if **A** is a finite field then $x_n \in CCD(\mathbf{A})$. This is the final conclusion of Theorem 3.

Given that q_n is essentially measurable on $\{x_{n-d}, q_{n-1}, \cdots, q_{n-m}\}$, for each n, we conclude by induction that q_n is essentially measurable $\{x_{n-d}, x_{n-d-1}, q_{n-2}, \cdots, q_{n-m-1}\}$, $\cdots$ and finally then that q_n is essentially measurable $\{x_{n-d}, x_{n-d-1}, \cdots\}$. Define

$$q'_n = q_{n+d},$$

$$y'_n = y_{n+d}, \qquad n = 0, \pm 1, \cdots .$$

Then it is a simple translation of notation to verify that $\{x_n, q'_n, y'_n\}$ is a Q-coded communication system with decoder memory span m that is $\{\psi'_n, 0, m\}$-optimal, where $\psi'_n = \psi_{n+d}$, $n = 0, \pm 1, \cdots$. This is the second conclusion of Theorem 3.

Finally, if $d = 0$, then "$\{x_n\} \in CCDf$" may be replaced by: "$\{x_n\} \in CCD$." Then M is unrestricted, since no "future" is involved that must be restricted to a finite field. This completes the proof.

Theorem 2 is a limiting case of Theorem 3, proved by putting $m = \infty$ everywhere in the proof of Theorem 3.

V. A COROLLARY

It is a consequence of Lemma 2 and of the proof of Theorem 4 that, given ω, in a set of probability one, $q(\omega)$ is that unique value of p which minimizes $\psi(x(\omega), z_p(\omega))$. (This was remarked in connection with equation 25.) Applying this to the situation of Theorem 1, one sees that the transmitter of a delay-free Q-coded communication system $\{x_n, q_n, y_n\}$ satisfying Theorem 1 has the block diagram form shown in Fig. 1. (If $d > 0$, one simply puts an analog delay line in the input lead, ahead of the rest of the system.)

This block diagram can be described thus: at time subsequent to $t = n - 1$ and prior to $t = n$, the transmitter has in its digital store the values $q_{n-1}, q_{n-2}, \cdots$ of the previously transmitted signals. From these, quantities $z_{1,n}, z_{2,n}, z_{3,n}, \cdots$ are constructed. These are the z_p of Lemma 2, for the particular random variable y_n. When x_n becomes available, quantities $\psi_n(x_n, z_{1,n})$, $\psi_n(x_n, z_{2,n})$, $\cdots$ are constructed and the comparator identifies the least of these (unique with probability one). The transmitted q_n is that value of the index which identifies the least $\psi_n(x_n, z_{p,n})$. This index is transmitted to the receiver as q_n and is also stored in the transmitter's memory for the next cycle. The receiver can be realized using a portion of the transmitter, as suggested in Fig. 2. Each function generator in these diagrams can

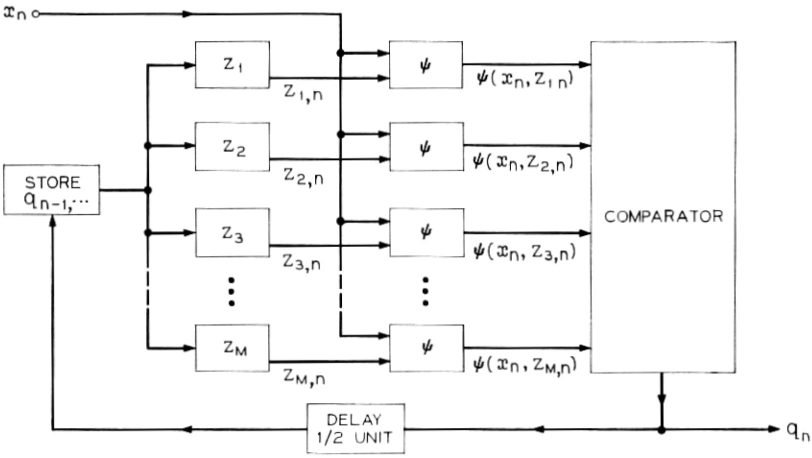

Fig. 1 — Generalized form of optimum transmitter.

of course be nonstationary. Connections to a master "clock" are not shown.

VI. REMARKS ON K AND CCD

One might ask to what degree are the central hypotheses of Theorem 4 necessary to the conclusions. The theorem itself provides a partial answer: conclusions *i* and *ii* do not use $x \in CCD(\mathbf{G})$ at all, and use only a measurability and a boundedness property of ψ. The critical conclusions are the uniqueness conclusions *iii* and *iv*. Clearly, something

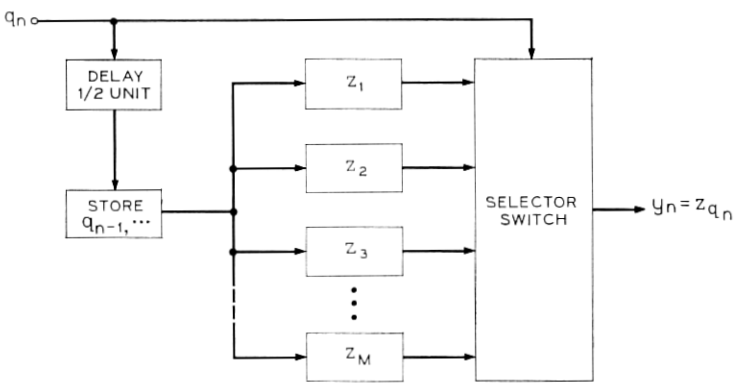

Fig. 2 — Form of receiver.

is required of $\psi(x, y)$ that makes it, in some sense, smaller when $y = x$ than elsewhere, and not too indifferent to the value of y when $y \neq x$, if uniqueness is to be expected from the hypothesis of ψ-optimality. As we have already noted, the hypothesis $\psi \varepsilon K$ is fairly weak in this regard, and could, in the presence of CCD, be made weaker at the expense of further elaboration of the proof.

The interesting hypothesis is $x \varepsilon CCD(\mathbf{G})$. This implies that if x has a conditional probability distribution relative to the field $\mathbf{G}$, then that distribution is continuous. It is easy to see that the ψ-optimum quantizing of a random variable x need not be unique if the distribution of x is not continuous, even when one uses $\psi(x, y) = (x - y)^2$. Since y in Theorem 2 ψ-optimally quantizes x for each event measurable on the conditioning field $\mathbf{G}$, something like $x \varepsilon CCD(\mathbf{G})$ is necessary if conclusion *iv* is to follow. Thus we conclude a loose kind of necessity for this hypothesis.

We notice finally that *iii* and *iv* were proved by confining the redundancy among the $\{z_p, p \varepsilon Q\}$ to a null set. In the application of this idea to the situation of Theorem 1, it seems likely that redundancy in the $\{z_{pn}, p \varepsilon Q\}$ for some fixed n might indeed be exploited to improve some

$$E\{\psi_{n+k}(x_{n+k}, y_{n+k})\}, \qquad k > 0, \tag{29}$$

by selection, among the minimizing z_{pn} to which $E\{\psi_n(x_n, y_n)\}$ is indifferent, one which actually contributes information about x_{n+k} and therefore allows a reduction in (29). I have no example to show this phenomenon, so its existence remains a conjecture. We have proved, of course, that its possible existence is ruled out by $x \varepsilon CCD(\mathbf{G})$.

REFERENCES

1. Shannon, C. E., "A Mathematical Theory of Communication," B.S.T.J., *27*, Nos. 3 and 4 (July and October 1948), pp. 379–423, 623–656.