MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES
3 4456 0361266 9

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION

4

ORNL-2533
Health and Safety

EVIDENCE FOR THE UNIFORM RADIATION LIFE SHORTENING PREDICTED BY INFORMATION THEORY

J. S. Mendell

OAK RIDGE NATIONAL LABORATORY
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION
LIBRARY LOAN COPY
DO NOT TRANSFER TO ANOTHER PERSON
If you wish someone else to see this document, send in name with document and the library will arrange a loan.

Printed in USA. Price $2.00 cents. Available from the

Office of Technical Services
U. S. Department of Commerce
Washington 25, D. C.

LEGAL NOTICE

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, express or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission to the extent that such employee or contractor prepares, handles or distributes, or provides access to, any information pursuant to his employment or contract with the Commission.

ORNL-2533

Contract No. W-7405-eng-26

HEALTH PHYSICS DIVISION

EVIDENCE FOR THE UNIFORM RADIATION LIFE SHORTENING PREDICTED BY INFORMATION THEORY

J. S. Mendell

DATE ISSUED

JUN 11 1958

Submitted as a Thesis to the Graduate School of Vanderbilt University in partial fulfillment of the requirements for the degree of Master of Arts in Physics

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee
operated by
UNION CARBIDE CORPORATION
for the
U. S. ATOMIC ENERGY COMMISSION

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES

3 4456 0361266 9

ii

ACKNOWLEDGMENTS

This research was suggested and supervised by Dr. H. P. Yockey.

The Operation Greenhouse data were furnished by Dr. A. C. Upton.

Illustration 2 is used with the permission of Dr. R. K. Mortimer. Illustration 3 is used with the permission of Dr. Thomas Wood.

The author wishes to thank Dr. Allyn Kimball and Dr. Walter Snyder for many helpful conversations.

This research was performed at Oak Ridge National Laboratory, Health Physics Division, while the author was an AEC Radiological Physics Fellow. The Fellowship was administered by the Oak Ridge Institute of Nuclear Studies.

The author wishes to thank his fellowship advisers, Dr. E. E. Anderson and Dr. J. I. Hopkins.

TABLE OF CONTENTS

	Page
Acknowledgment	ii
List of Tables	v
List of Figures	vi
Chapter	
I. Survey of Information Theory	1
A. Prefatory Remarks	1
B. Information	1
C. The Average Information Per Symbol: Information Entropy	2
D. Conditional Entropy, Read-Off Error, and Noise	4
E. Redundancy	4
F. More About Information and Noise	5
G. Genetic Information	5
H. Biological Redundancy	6
I. Biological Noise	7
II. The Mathematics of Information Theory	9
A. Definition of Information	9
B. Average Information Per Symbol	9
C. A Lemma	10
D. Joint Information and Conditional Entropy	11
E. A Measure of Noise	13
III. Yockey's Contributions	15
A. Introduction	15
B. A Genetic Alphabet	15

TABLE OF CONTENTS

Chapter Page

C. The Measure of Biological Noise - - - - - - - 16

D. Accumulation vs. Elimination of Read-Off Error 16

E. Equivocation - - - - - - - - - - - - - - - - - 17

F. Ensembles of Organisms - - - - - - - - - - - 18

G. The Entropy Probability Distribution - - - - 20

H. Time Dependence - - - - - - - - - - - - - - - - 22

I. Explicit Time Dependence of H and $dH/d\lambda$ - - - 24

J. The Shape of Survival Curves - - - - - - - - 24

IV. Analysis: Procedure and Results - - - - - - - - 32

A. Prefatory Remarks - - - - - - - - - - - - - - 32

B. An Hypothesis - - - - - - - - - - - - - - - - 32

C. An Ideal Experiment - - - - - - - - - - - - - 32

D. Operation Greenhouse - - - - - - - - - - - - 33

E. Explanation of Tables - - - - - - - - - - - - 35

F. Consistency of Values - - - - - - - - - - - - 35

G. Evaluation of Results - - - - - - - - - - - - 36

V. Discussion - - - - - - - - - - - - - - - - - - - 43

Appendix I

Moments and Their Functions - - - - - - - - - - - - - 44

Appendix II

A Derived Entropy Distribution - - - - - - - - - - - 49

Appendix III

Explicit λ Dependence of H and $dH/d\lambda$ - - - - - - - 51

Bibliography - 58

LIST OF TABLES

Table		Page
1.	$\bar{x}$, σ, β_1, β_2 for Greenhouse Gamma Rays.	37
2.	$\bar{x}$ and σ - Greenhouse Neutrons - Easy and George Shots	38
3.	$\bar{x}$ and σ - Greenhouse Laboratory X-Rays	39
4.	Z and p(Z) for σ - Operation Greenhouse - Gamma Rays	40
5.	Z and p(Z) for β_1 - Operation Greenhouse - Gamma Rays	41
6.	Z and p(Z) for β_2 - Operation Greenhouse - Gamma Rays	42
7.	Specimen Calculation of $\bar{x}$, σ, β_1 and β_2	46
8.	Specimen Calculation of $\bar{x}$, σ, β_1 and β_2 (continued)	47
9.	Specimen Calculation of $\bar{x}$, σ, β_1 and β_2 (concluded)	48

LIST OF ILLUSTRATIONS

Figure		Page
I.	A Time-Dependent Entropy Distribution	20
II.	Percentage Survival of Yeast (Mortimer, 1956) . .	25
III.	Thermal and X-Ray Killing of Yeast (Wood; 1954, 1956)	26
IV.	Increment of Area About to Cross H_d	27

I. SURVEY OF INFORMATION THEORY

Prefatory Remarks. This chapter is a guide to Chapters II and III, containing the basic vocabulary, and outlining the important ideas without their algebraic derivation.

Information theory was developed by electrical engineers to discover the most efficient way of expressing a message for use in transmission systems. A simple example of this efficient coding is the use in Morse code of the short di-dah symbol for the often-used "A" and the long dah-dah-di-dah and dah-di-di-dah for the seldom-used "Q" and "X." Morse code, of course, was developed long before information theory, as it was formulated by Shannon (1949).

Information theory is useful because it shows how to transmit in a communications channel between a source and a receiver in the presence of noise, and because it establishes rules for working within the capacity of the system.

Information. Suppose an experimenter searches for Newton's Optiks in a 5,000-volume library. His first random selection has an a priori probability of 1/5000 of yielding the book. His use of the card index will increase this probability to unity. The card catalog has yielded 12.3 bits of information. Information (in bits or binary units) is defined here as

$$\log_2 \frac{1}{1/5000} .$$

2

Why use the definition $\log_2 \frac{P_{final}}{P_{initial}}$ for information instead of $\frac{P_{final}}{P_{initial}}$?

Suppose the experimenter searches for both Optiks and Principia. The initial probability will be $(\frac{1}{5000} \cdot \frac{1}{5000})$, the final probability will be unity, and the information 24.6 bits. Thus, the logarithmic definition doubles the information (which seems plausible) although the probability has been squared.

This definition is appealing, because it causes the information of successive choices to be the sum of information of individual choices, because it bases the theory on experimentally determined or theoretically predicted probabilities, and because it leads directly to a successful theory.

The Average Information per Symbol: Information Entropy. Theorists are principally concerned with the average information/symbol, which is the density of information in a message. This is because a highly informative short message is usually more economical than an equally informative long message. The average information per unit time is the product of average information/symbol and average number of symbols/time. When the symbol generating capacity of a transmitter has been maximized, it is necessary to increase the average information content of the symbols.

There is a formula for the average information/symbol in a long message.

Average information/symbol $= H(x) = -\sum_i p(i) \log_2 p(i)$ bits/symbol.

Here p(i) represents the probability that the i-th type of symbol will appear in any position x. It is the fraction of the message symbols that are of the i-th type.

The factor $-\log_2 p(i)$ is the information of a symbol with probability p(i). Its coefficient p(i) is then a weighting factor. More simply, an infrequent letter has a high information value (small p(i); large $-\log_2 p(i)$) but is penalized for its infrequent appearance in the message.

The statistical thermodynamic formula for entropy is

$$s = -k \sum p \log p,$$ where k is Boltzman's constant,

so the average information/symbol is named by analogy the <u>information entropy</u>.

It must be clear from the library problem that information is a measure of the uncertainty abolished by a choice, or briefly a measure of the uncertainty, or a measure of choice.

If a system can exist in only one state, its information entropy is zero, as may be verified by substituting

$$p(1) = 1,\ p(2) = 0,\ p(3) = 0, \text{ etc.}$$

in the equation for entropy.

If a system exists in one of many equally, or almost equally, possible states, it possesses a high degree of choice, and a high information entropy.

Later we will say that certain biological systems possess high information entropy, and we will mean that they have high choice in the synthesis of proteins.

Conditional Entropy, Read-Off Error and Noise. The function conditional information entropy is used as a measure of noise. Conditional information entropy is defined in terms of conditional probabilities.

$$H_x(y) = -\sum_{i,j} p(i)\ p_i(j)\ \log_2\ p_i(j)$$

Here $p_i(j)$ is the probability of a one-to-one correspondence between the j-th symbol in place Y and some preassigned i-th symbol in place X. For example, $p_Q(X)$ is the probability that the transmitted Morse code "Q" (dah-dah-di-dah) will be erroneously read-off at the receiver as "X" (dah-di-di-dah). There are $(28)^2$ $p_i(j)$ for Morse code, corresponding to $(28)^2$ one-to-one correspondences between transmitted and received signals. Read-off error is the name for the $(28)^2 - 28$ types of incorrect transmission. Read-off error is the manifestation of noise, which is, roughly speaking, a nonsense message. If there were no noise, there would be no read-off error, and

$$H_x(y) = 0$$

Redundancy. Redundancy is a partial or complete repetition of message content which often allows individual errors to be detected and corrected.

For instance, the redundant message HELD COST A SEAT KELP LOFT AT SEA may be correct to read HELP LOST AT SEA.

One might suppose that is impossible to transmit information in the presence of noise, or at least in the presence of considerable noise. (Noise is the general term applied to all the phenomena which prevent correct transmission: static, oscillations, stray currents, inattention,

etc. Information theory deals with the effects of noise generated by stochastic processes.) But, in fact, it has been shown possible, both in theory (Shannon, 1949) and practice to transmit in the presence of noise. By employing sufficient redundance it is possible to transmit enough information to reconstruct even a badly altered message. The greater the noise, the greater the required redundance, but transmission is still possible.

More about Information and Noise. Information entropy is a measure of the uncertainty of a system. That is, if a system exists in one of many possible states, it possesses high uncertainty. Any description of the system eliminates some or all of the uncertainty concerning the state of the system, hence contains high information content.

Information entropy is a measure of available choice. If a system possesses high uncertainty, hence high information entropy, it may exist in any one of many states. Hence, the choice of one of these states is highly significant.

Any process which eliminates choice, or decreases uncertainty, destroys information. For instance, a stochastic process, which acts upon the elements of a system, reduces choice, for a stochastic process cannot be controlled. Hence, the conditional entropy, which is a measure of the entropy of the stochastic process, must be subtracted from the information entropy. This will be expressed mathematically in Chapter II.

Genetic Information. That the nuclear material is responsible for heredity was established by the fact that in sexual reproduction the male gamete consists almost entirely of nuclear material. Heredity has been

traced down beyond the chromosomes, through the genes, to twenty amino acid residues. Moreover, there is a strong possibility that the exact order of four kinds of nucleotides is the molecular vehicle carrying genetic information (Watson and Crick, 1955).

Thus every cell, from our point of view, contains a communications system in which specificity is carried from the genome to the microsomes in the cytoplasm by means of ribose nucleic acid. There are genetic counterparts of an alphabet, noise, and redundancy.

The alphabet is, of course, a list of all symbols used in the genetic message. In the genome either the twenty amino acid residues or the four nucleotides guanine, adenine, cytosine, and thymine may comprise the alphabet. Since there is strong evidence that the order of four types of nucleotides determines the protein specificity (Watson and Crick, 1953) this paper will provisionally assume that the primary alphabet is composed of four letters.

Gamow and Ycas (1958; also Gamow, Ycas, and Rich, 1956) have investigated the code between the four letter nucleotide alphabet and the twenty letter amino acid residue alphabet.

Here we will describe Yockey's investigations of the effect of aging and deleterious agents on the nucleotide symbols.

Biological Redundancy. Repetition exists in all multicellular organisms derived from a single cell, for the genome of each cell carries the same information as that of its neighbor. Moreover, diploid organisms carry alternate messages in their duplicate chromosomes. There seem to be other types of redundancy, not necessarily associated with the duplication of some large nuclear structure.

It was pointed out previously that messages may be transmitted in the presence of noise, provided sufficient redundancy is incorporated in the message. We will see that, from an information theory viewpoint, cells may function in the presence of biological noise (defined below) because considerable redundancy is inherent in the genetic transmitter. We will also see that when the noise exceeds the redundancy of the genetic message, the cell must cease to function.

Biological Noise. Noise is the aggregation of phenomena which prevent perfect transmission of messages.

The biological noise most easily treated by information theory is the type which causes one symbol to mime another. Radiation, thermal agitation, and biochemical side reactions, which are representable by stochastic processes, may all raise nucleotides to excited states and cause read-off error. But this noise information, i.e., conditional information entropy, decreases choice. The read-off error generated by noise must be detected and corrected by the use of redundant information. Hence, noise reduces the net useful information entropy of the cell. Unless the nucleotides recover from their unusual state read-off error accumulates. Later we will identify aging with the accumulation of read-off error in the genome.

Read-off error may accumulate within the genome without destroying the cell's ability to function, providing it does not exceed the redundant capacity of the cell. This is equivalent to saying that a cell may resist certain amounts of radiation. It is not, however, equivalent to the statement that the cell is not harmed by radiation; since the read-off error accumulates, the next radiation dose, or eventually natural aging, will be additive and may cause read-off error to exceed the redundant capacity.

If, as we believe, genomes differ in their redundant capacity then they will appear to differ in radiosensitivity.

II. THE MATHEMATICS OF INFORMATION THEORY

Definition of Information. Information is a measure of uncertainty or a measure of choice.

Mathematically,

$$\text{information} = \log_2 \frac{p_f}{p_i} \qquad \text{(Eq. 1)}$$

where p_i and p_f are the initial and final probabilities of the occurence of some event. This definition displays information as a measure of change of uncertainty, or of freedom of choice.

Average Information Per Symbol. Brillouin (1956) derives the average information per symbol in this way. Consider M different types of symbols, 1, 2, 3, . . . j, . . . M, occurring with frequency N_1, N_2, . . . N_j, . . . , N_M and with probabilities $p(1) = N_1/G$, $p(2) = N_2/G$, . . . $p(j) = N_j/G$, . . . , $p(M) = N_M/G$, where $G = \sum_j p(j)$. These are to be distributed at random among G cells, one symbol per cell.

Each possible arrangement is a message. The number of messages is

$$\frac{G!}{\prod_j N_j!}$$

The probability of one particular message is then

$$P = \frac{\prod_j N_j!}{G!} \qquad \text{(Eq. 2)}$$

The information of a message is

$$\begin{aligned} \text{information} &= -\log_2 \prod N_j! + \log_2 G! \qquad \text{(Eq. 3)} \\ &= -\log_2 N_j! + \log_2 G! \end{aligned}$$

By Stirling's approximation

$$\ln 2 \log_2 \quad Q! = Q(\ln Q - 1) \qquad \text{for } Q > 100$$

$$\log_2 \quad Q! = \frac{Q}{\ln 2} (\ln Q - 1)$$

$$\text{information} = -\sum \frac{N_j}{\ln 2} (\log_e N_j - 1) + \frac{G}{\ln 2} (\log_e G - 1)$$

$$= -\sum N_j \log_2 N_j + G \log_2 G + \sum \frac{N_j}{\ln 2} - \frac{G}{\ln 2}$$

$$= -\sum_j N_j \log_2 N_j + G \log_2 G + \frac{1}{\ln 2} \left[\sum_j N_j - G \right]$$

Since $G = \sum_j N_j$

$$\text{information} = G \log_2 G - \sum N_j \log_2 N_j$$

$$\text{information} = -G \sum_j \frac{N_j}{G} \log_2 \frac{N_j}{G} = -G \sum_{j=1}^{n} p(j) \log_2 p(j)$$

Hence, the information/symbol of an alphabet is

$$\text{information}/G = -\sum p(i) \log_2 p(i)$$

This function is named the information entropy, H(x).

$$H(x) = -\sum p(i) \log_2 p(i) \quad \text{bits/symbol} \qquad \text{(Eq. 4)}$$

<u>A Lemma</u>. We will prove the lemma

$$-\sum_i p(i) \log_2 q(i) \geq -\sum_i p(i) \log_2 p(i)$$

where p(1), p(2), . . . , p(i), . . . , p(n) is one set of probabilities, $\sum_{i=1}^{n} p(i) = 1$, and q(1), q(2), . . . , q(i), . . . q(n) is another set of probabilities, $\sum_{i=1}^{n} q(i) = 1$. That is, q(1), . . . , q(n) is any set of n positive fractions whose sum is unity.

$$\sum p(i) = \sum q(i) = 1$$

$$q(i) = p(i) + u(i) = p(i) \left[1 + \frac{u(i)}{p(i)} \right] \qquad \text{(Eq. 5)}$$

$$\sum u(i) = \sum q(i) - \sum p(i) = 0 \qquad \text{(Eq. 6)}$$

Define $\psi = \sum p(i) \log_2 q(i)$ (Eq. 7)

$$= \sum p(i) \log_2 p(i) \left(1 + \frac{u(i)}{p(i)}\right)$$

$$= \sum p(i) \log_2 p(i) + \sum p(i) \log_2 \left(1 + \frac{u_i}{p_i}\right)$$

Now $\frac{u}{p} \geq \log_2 \left(1 + \frac{u}{p}\right)$

where both functions are defined so

$$\sum p(i) \log_2 p(i) + \sum p(i) \log_2 \left(+ \frac{u(i)}{p(i)}\right) \leq$$

$$\sum p(i) \log_2 p(i) + \sum p(i) \frac{u(i)}{p(i)}$$

$$\sum p(i) \log_2 p(i) \leq \sum p(i) \log_2 p(i)$$

$$-\sum p(i) \log_2 q(i) \geq -\sum p(i) \log_2 p(i) \qquad \text{(Eq. 8)}$$

<u>Joint Information and Conditional Entropy</u>. Conditional entropy was introduced in Chapter I as a measure of noise. This function is constructed of the ordinary $p(i)$ and two new probabilities, $p(i,j)$ and $p_i(j)$.

The <u>joint probability</u>, $p(i,j)$, is the probability of the occurrence of the i-th and j-th types of symbols in positions X and Y.

The conditional probability, $p_i(j)$, is the probability of the successive occurrence of two symbols in positions X and Y.

Thus, $p(i,j)$ represents occurrence of symbols of the i-th and j-th types, occurrence of the second being either independent or dependent of

the outcome of the occurrence of the first. But $p_i(j)$ represents dependent occurrence of symbols of the i-th and j-th type, occurrence of the second being dependent on the outcome of the occurrence of the first.

$$\sum_i \sum_j p(i,j) = \sum_{i,j} p(i,j) = 1 \qquad \text{(Eq. 9)}$$

$$p(i) = \sum_j p(i,j)$$
$$p(j) = \sum_i p(i,j) \qquad \text{(Eq. 10)}$$

$$\sum_{i,j} p(i) \cdot p(y) = 1 \qquad \text{(Eq. 11)}$$

$$p(i) \cdot p_i(j) = p(i,j) \qquad \text{(eq. 12)}$$

The <u>joint information</u> entropy is

$$H(x,y) = -\sum_{i,j} p(i,j) \log_2 p(i,j) \qquad \text{(Eq. 13)}$$

while the informations for x alone and y alone are

$$H(x) = -\sum_i p(i) \log_2 p_i(j) = -\sum_{i,j} p(i,j) \log_2 p(i)$$

$$H(y) = -\sum_{i,j} p(i,j) \log_2 p(j)$$

Moreover

$$H(x) + H(y) = -\sum_{i,j} p(i,j) \log_2 p(i) \cdot p(j) \qquad \text{(Eq. 14)}$$

Comparing

$$-\sum_{i,j} p(i,j) \log_2 p(i) \cdot p(j)$$

with $-\sum p(i,j) \log_2 p(i,j)$

and remembering the lemma (Eq. 8), we find that

$$H(x) + H(y) \geq H(x,y) \qquad \text{(Eq. 15)}$$

The conditional entropy, $H_x(y) = -\sum_{i,j} p(i)\ p_i(j) \log_2 p_i(j)$ (Eq. 16) is the weighted average of the entropy of successive occurrences in positions x and y.

The conditional entropy of the occurrence of any symbol j in position y, following the occurrence of some particular symbol i in position x is

$$H_i = -\sum_j p_i(j) \log_2 p_i(j)$$

The total conditional entropy for all possible symbols in position x is the summation over all possible i, weighted according to the probability of each i:

$$H_x(y) = -\sum_{i,j} p(i)\ p_i(j) \log_2 p_i(j) \qquad \text{(Eq. 17)}$$

We have shown that (Eq. 15)

$$H(x) + H(y) \geq H(x,y)$$

We now show that

$$H(x) + H_x(y) = H(x,y) \qquad \text{(Eq. 18)}$$

For $$H(x) + H_x(y) = -\sum_{i,j} p(i,j) \log_2 p(i) - \sum_{i,j} p(i,j) \log_2 p_i(j)$$

$$= -\sum_{i,j} p(i,j) \log_2 p(i) \cdot p_i(j)$$

$$H(x) + H_x(y) = H(x,y)$$

It follows that

$$H(x) + H(y) \geq H(x) + H_x(y)$$

$$H(y) \geq H_x(y) \qquad \text{(Eq. 19)}$$

equality occurring when the occurrence of y is independent of the occurrence of x.

<u>A Measure of Noise</u>. If x is the transmitter, and y is the receiver, $p_i(j)$ is the probability of correct transmission of a symbol when i = j and

the probability of incorrect transmission when $i \neq j$. Then $H(x)$ is the information entropy of the source, and $H(y)$ is the information entropy of the receiver.

For a noiseless transmission $p_i(j) = 0$ for all $i \neq j$ so that $H_x(y) = 0$, $H(y) \geq 0$.

For a completely noisy transmission, the transmission has no effect on the reception, and

$$p(j) = p_i(j)$$

Shannon (1949) has proved that $H_x(y)$ is the proper measure of the deleterious effects of noise generated by stochastic processes. We have already explained that noise destroys uncertainty (or choice). Hence the choice that remains after a stochastic process is measured by

$$T = H(y) - H_x(y)$$

which we will call the net entropy.

This is equal to

$$T = H(y) + H(x) - H(x,y)$$

or, by Eq. 13 and Eq. 14

$$T = -\sum_{i,j} p(i,j) \log_2 p(i) \cdot p(j) + \sum_{i,j} p(i,j) \log_2 p(i,j)$$

Hence $T \geq 0$ always.

In particular, if $p(i,j) = p(i) \cdot p(j)$, $p_i(j) = p(j)$ and $T = 0$.

Thus $T = 0$ for the completely noisy case.

III. YOCKEY'S CONTRIBUTIONS

Introduction. In this chapter we will show that information entropy, conditional entropy, read-off error, and other information theoretical ideas have counterparts in genetic systems.

In order to present an undistorted picture of Yockey's work, we will present more than is necessary for our own calculations, but we will not attempt to justify hypotheses which do not provide the background for our own work.

A Genetic Alphabet. Certain biochemicals occur universally in nucleic structures. Yockey looked to these for his genetic alphabet, applying information theory to the Watson-Crick DNA model. Although the genetic alphabet may be identified with the twenty amino acid residues which exist in proteins (Gamow, Ycas and Rich, 1956; Gamow and Ycas, 1958), Watson and Crick (1953) have suggested that the four nucleotides which occur in nucleic acid constitute the genetic alphabet.

Yockey's theory does not depend on a deep understanding of the behavior of these nucleotides, but on the knowledge that there are four, that they occur in various orders (Watson and Crick, 1953), and that they may exist in tautomeric states (Yockey, 1956).

The Watson and Crick model gives concrete meaning to the idea of genetic words. According to this model, a DNA molecule (the word) is a double helix with nucleotides (the symbols) located at various sites. From our viewpoint different nucleotide orders spell out different words.

The Measure of Biological Noise. Yockey postulates that the four nucleotides ordinarily exist in unexcited states S_A, S_B, S_C, S_D, with probabilities p(A), p(B), p(C), p(D); $\sum_i p(i) = 1$. When the nucleotides are excited by ambient energy (from thermal fluctuations, chemical side reactions, radiation, etc.) they are apt to assume tautomeric states, and the probability of the transition $S_i \longrightarrow S_i^*$ is clearly a conditional probability. Moreover, he assumes that the tautomeric state often mimes an unexcited state. That is, its symbol imitates one of the three normal symbols, so that $S_i^* \backsim S_j$. This results in an incorrect transmission, the manifestation of noise, whose measure is the contitional entropy

$$H_x(y) = -\sum_{i,j} p(i)\, p_i(j) \log_2 p_i(j)$$

But $p_i(j)$ is the probability of the process $S_i \longrightarrow S_j$, and the function is defined for thermal-, chemical- and radiation-induced noise, or, more precisely, for noise generated by some stochastic process which approximates them.

Accumulation vs. Elimination of Read-Off Error. Read-off error must accumulate in the cells, for tautomeric forms cannot return to ground state until raised to still higher energy states.

Since thermal agitation, chemical side reactions, and radiation are ubiquitous and more apt to produce error than correct it, read-off error must increase with the age of a nucleus. Consequently biological communications systems become noisy and fail to direct normal cell functions such as metabolism, reproduction, and recovery from injury. Radiation above background can only aggravate the noisiness of the channel. In this sense radiation is equivalent to accelerated aging.

Aging seems to occur in both the somatic cells and in the germ cells. A number of experiments bear on the two competing processes of accumulation and elimination of error from the germ line. Lansing (1947, 1948) demonstrated that the viability of rotifer offspring depends on the age of their parents (hence, the age of their parent's germ cells). Hatching of eggs laid late in the life of these parthogenetic animals leads to extinction within a few generations. But this trend is reversed by hatching of the eggs of the young daughters of old parents. Thus, artificial selection, combined with the natural germ-line error correction, restores vigor.

Equivocation. The number of symbols in its alphabet and their p(i) determine the maximum entropy of a system via Eq. 4. For the genetic system with its four symbol alphabet, if all $p(i) = \frac{1}{4}$, the maximum entropy is 2 bits/symbol. However, read-off error reduces choice and conditional entropy, which is called equivocation by Shannon (1949), must be subtracted from the maximum entropy. We will later display Yockey's interesting proof that all the members of an ensemble must accumulate appreciable equivocation.

Now it is clear that organisms may sustain a certain amount of equivocation. X-ray induced mutations provide evidence for this. Nevertheless, organisms must die, or at least cease to function, when their equivocation becomes too great. In other words, organisms do not require 2 bits/symbol of entropy to survive and function; some of the 2 bits/symbol is associated with redundancy.

We stated earlier that cells may continue to function in the presence of noise until the accumulated noise exceeds the inherent redundancy of the genome. According to theory, the number of uncorrected errors may be made as small as desired by using sufficient redundancy. Thus, the amount of redundancy

required depends not only on the amount of accumulated noise, but also on the number of uncorrected errors that are tolerable.

If a cell must survive in an extremely hostile environment, it cannot tolerate many errors. Hence, it cannot afford to lose as much entropy as a cell functioning in an ideal environment. We say then that a cell possesses a lower entropy limit, H_d, which is environment dependent.

The idea may be extended to the organism constructed of the individual cells, in which case H_d is an average value. Incidentally, "death" has an indefinite meaning. In bacteria death is the inability to reproduce; in mammals death is a sharp decrease in body temperature.

Biologists are often puzzled by the abrupt death of an apparently healthy animal. This phenomenon may be interpreted from an information theoretical viewpoint.

Initially the animal's communications system must possess some equivocation. This equivocation is counteracted by the redundancy inherent in the system, and the animal is healthy. With increasing age, the equivocation increases, but is counteracted by correspondingly more redundance. Thus the animal displays no signs of weakness, for increasing equivocation never exceeds the correction capacity of the system. Eventually, however, the animal's net entropy falls below H_d, i.e., equivocation abruptly exceeds redundancy, and the animal dies.

Ensembles of Organisms. Since his theory is statistical, Yockey introduces the statistical mechanical idea of ensembles of organisms. He shows that although members of an ensemble may be isogenic, they must vary in net genetic entropy content. That is, members of an isogenic ensemble have suffered, at any time, various amounts of equivocation. Later we will show that organisms afflicted at birth with the same equivocation remain equal in

equivocation until they die - simultaneously. This, and the experimental fact that offspring of the same parents die at different ages, establishes that equivocation is distributed unequally at birth, and remains unequally distributed. Of course, this is equivalent to a distribution of net entropy.

Yockey has shown by the following reductio ad absurdum proof that equivocation must become appreciable in the genome.

Net message entropy is $H = H_o - H_x(y)$, where H_o is the initial entropy and $H_x(y) = \sum_{i,j} p(i)\, p_i(j) \log_2 p_i(j)$, the equivocation. The derivative with respect to λ, some measure of time or other deleterious agent, is

$$\frac{dH}{d\lambda} = \frac{1}{\ln 2} \sum_{i,j} \Big[p(i) \frac{d}{d\lambda} p_i(j) + p(i) \frac{d}{d\lambda} p_i(j) \ln p_i(j) + p_i(j) \frac{d}{d\lambda} p(i) \ln p_i(j) \Big] \qquad \text{(Eq. 20)}$$

Some read-off error must always occur, for background radiation, thermal agitation, and chemical side reactions are ubiquitous. In fact, there is evidence that both somatic and genetic mutations occur. Hence, $\frac{dp_i(j)}{d\lambda} \neq 0$. Moreover, $p(i) \neq 0$.

Nevertheless, it might be supposed that $p_i(j)$ for $i \neq j$ never becomes large enough to cause appreciable loss of net entropy. In other words, that $p_i(j) \longrightarrow 1$ for $i = j$, $p_i(j) \longrightarrow 0$ for all $i \neq j$, and $\frac{dH}{d\lambda} = 0$.

But consider the term $p(i)\frac{d}{d\lambda} p_i(j) \ln p_i(j)$ in Eq.20. The $\ln p_i(j)$ will become infinite for the case $i \neq j$ and its coefficient $p(i) \frac{d}{d\lambda} p_i(j)$ is non-zero. Thus, for the errorless case $\frac{dH}{d\lambda} \longrightarrow -\infty$ instead of zero.

Hence, enough error must accumulate in all genetical messages to cause an appreciable $\frac{dH}{d\lambda}$.

The Entropy Probability Distribution. Since the net entropy of each member of an ensemble is time-dependent, the distribution of entropy is itself time-dependent. Such a function for a very large ensemble is shown below (Figure I); it is almost continuous.

First, the distribution is bounded on the left by $H = H_d$, and on the right by $H = H_{max} \leq 2$ bits/symbol.

Second, as λ increases the function moves leftward, perhaps changing shape.

UNCLASSIFIED
ORNL LR DWG. 29395

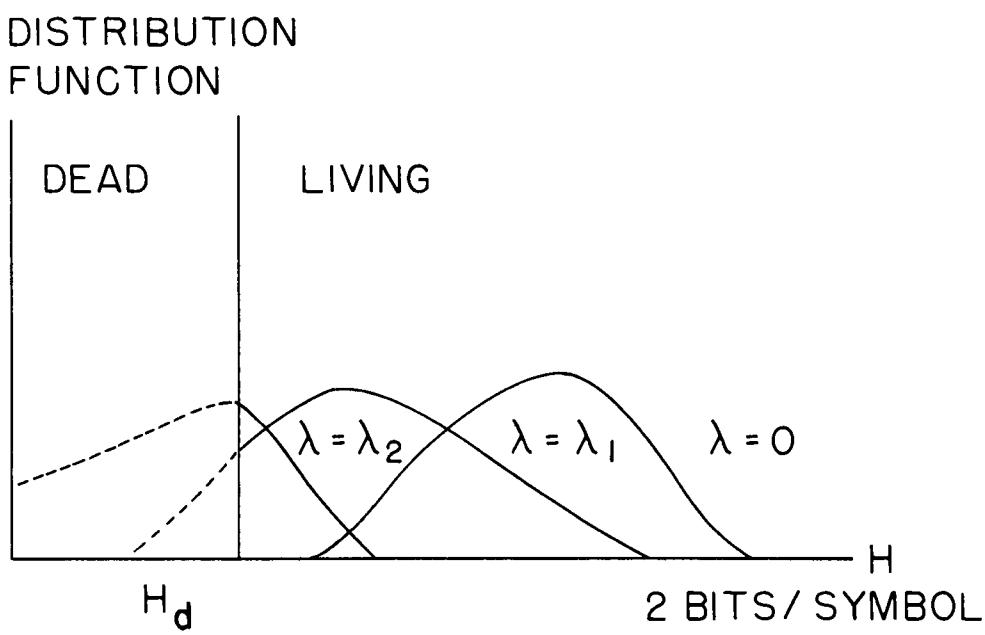

Figure I. A Time-Dependent Entropy Distribution

We have already explained that the lower entropy limit is set by the minimum entropy required for survival in any given environment. In Figure I the vertical line $H = H_d$ is a boundary between the organisms in the ensemble and those that have left it. The upper limit is, of course, about 2 bits per symbol.

It is easy to see why the curve moves leftward. Every point on the curve corresponds to the entropy of one or more organisms. As the organisms lose entropy their points shift to lower H values. If all animals in the ensemble lose entropy at the same rate, the curve will suffer a displacement; if they lose entropy at different rates, the curve will be distorted as well.

Moreover, as the curve shifts leftward its left edge crosses H_d and a fraction of it leaves the ensemble. This represents the death of a fraction of the ensemble.

Thus, aging is represented by the continuous leftward movement of the curve, and death as the passage of the curve across the vertical boundary H_d.

Yockey's distribution function is the probability distribution $\rho(H,\lambda)$. This function of H and λ is defined by

$$\frac{1}{\ell}\frac{d\ell}{d\lambda} = \rho(H_d,\lambda)\left.\frac{dH}{d\lambda}\right|_{H_d} \qquad \text{(Eq. 21)}$$

Here ℓ is the number of living organisms in the ensemble, so that $\frac{1}{\ell}\frac{d\ell}{d\lambda}$ is the rate of change of probability of survival. Since $\frac{dH}{d\lambda}$ is the rate at which a point on the curve shifts leftward, $\left.\frac{dH}{d\lambda}\right|_{H_d}$ is the rate at which the curve crosses the boundary H_d.

<u>Time Dependence</u>. Since the $p_i(j)$ are explicitly time dependent,

$$H = H_0 + \sum_{i,j} p(i)\, p_i(j) \log_2 p_i(j)$$

is implicitly time dependent. Although many researchers must have recognized this implicit dependence, Yockey was the first to point it out in print.

Yockey represents the $S_i \longrightarrow S_j$ transition as a first order chemical rate reaction so that

$$\frac{d}{d\lambda} p_i(j) = -J_{i,j}(\lambda)\, p_i(j) + C_{ij} \qquad \text{(Eq. 22)}$$

He explains (1958):

> The first term represents the loss of nucleotides resposible for the $S_i \longrightarrow S_j$ transition. The second term represents the gain in nucleotides engaging in the (i,j) transition coming from other nucleotides altered by the deleterious agent.

He simplified the expression.

$$\sum_j \frac{d}{d\lambda} p_i(j) = \frac{d}{d\lambda} \sum p_i(j) = -\sum_j J_{ij}(\lambda)\, p_i(j) + \sum_{i,j} C_{ij}$$

$$0 = -\sum_j J_{ij}(\lambda)\, p_i(j) + \sum_j C_{ij}$$

$$0 = -\sum_{i,j} J_{ij}(\lambda)\, p_i(j) + \sum_{i,j} C_{ij}$$

Employing average values for J_{ij} and $C_{ij}(\lambda)$

$$0 = -J(\lambda) \sum_{i,j} p_i(j) + C(\lambda) \sum_{i,j} 1$$

$$0 = -4\, J(\lambda) + 16\, C(\lambda)$$

$$C(\lambda) = \frac{1}{4} J(\lambda)$$

$$\frac{dp_i(j)}{d\lambda} = -J(\lambda)\, p_i(j) + \frac{1}{4} J(\lambda) = -J(\lambda) \left[p_i(j) - \frac{1}{4}\right] \qquad \text{(Eq. 23)}$$

$J(\lambda)$ may be expanded in a power series.

$$J(\lambda) = J_0 + J_1\lambda + J_2\lambda^2 + \ldots$$

in which successive terms on the right represent first order, second order, third order, . . . actions of the deleterious agents.

For instance, for haploid organisms $J_0 \gg J_1\lambda, J_2\lambda^2, \ldots\ldots$ Thus, $J(\lambda)$ represents a first order action with small higher order components.

For diploid organisms $J_1\lambda \gg J_0, J_2\lambda^2, J_3\lambda^3, \ldots$ so that $J(\lambda)$ represents a second order action with small higher and lower order components.

Thus, the leading order of the action depends on the ploidy and not on the deleterious agent, so that

$$J(\lambda) \doteq J_0 \qquad \text{(haploid case)} \qquad \text{(Eq. 24)}$$

$$J(\lambda) \doteq J_1\lambda \qquad \text{(diploid case)}$$

The numerical values of J_0 and J_1 will, of course, be greater for radiation aging than for natural aging.

It is important to appreciate that $J(\lambda)$ and $\rho(H,\lambda)$ are the embodiments of uniformity and individuality. The physico-chemical radiation interaction takes place in all organisms of an ensemble in the same way. The genetics of the ensemble is represented by $\rho(H,\lambda)$; this function is a description of the fact that members of an ensemble (even an isogenic one) differ in net entropy (or in accumulated equivocation).

There is another way of saying this: The members of an ensemble react in the same way to radiation, on the molecular level, but appear to differ in radiosensitivity, because some are better endowed with genetic entropy than their fellow members.

-24

Explicit Time Dependence of H and $dH/d\lambda$. In Appendix III we derive the explicit time dependent forms of H and $\frac{dH}{d\lambda}$, starting with the integration of $p_i(j) = -J(\lambda)\left[p_i(j) - \frac{1}{4}\right]$.

In general

$$\frac{dH}{dZ} = \frac{3}{4} u \log_2 \frac{1 - u Z}{1 + 3u Z} \qquad \text{(Eq. 25)}$$

$$H = H_0 + \frac{3}{4}(1 - uZ) \log_2 (1-uZ) + \frac{1}{4}(1 + 3uZ) \log_2 \frac{1}{4}(1 + 3uZ) \qquad \text{(Eq. 26)}$$

where

$$z = \exp -\int J(\lambda) d\lambda$$

and

$$u = 1 - 4g, \; g \ll 1.$$

These substitutions illustrate that an organism's aging depends on $e^{-\int J(\lambda)\, d\lambda}$. The aging time scale of an haploid organism depends on $e^{-J_0\lambda}$; the aging time scale of a diploid organism depends on $\exp\left[-\frac{1}{2} J_1 \lambda^2\right]$.

Owen and Mortimer (1956) experimented on both haploid and diploid yeast. They found an exponential survival curve (e^{-kt}) for x-ray killing of haploid yeast, and a sigmoidal survival curve (e^{-kt^2}) for x-ray killing of diploid yeast (Figure II). Thus, yeast seems to live on an exponential or sigmoidal time scale depending on ploidy - just as suggested by theory.

Thermal killing of diploid yeast also produces sigmoidal survival curves (Figure III) (Wood, 1954, 1956). This is predicted by theory which says that since thermal killing and radiation produce the same nucleotide reactions they must produce survival curves of the same shape.

The Shape of Survival Curves. The probability distribution was defined by Eq. 20:

$$\frac{1}{S}\frac{dS}{d\lambda} = \frac{1}{\ell}\frac{d\ell}{d\lambda} = \rho(H_d, \lambda) \left.\frac{dH}{d\lambda}\right|_{H_d}$$

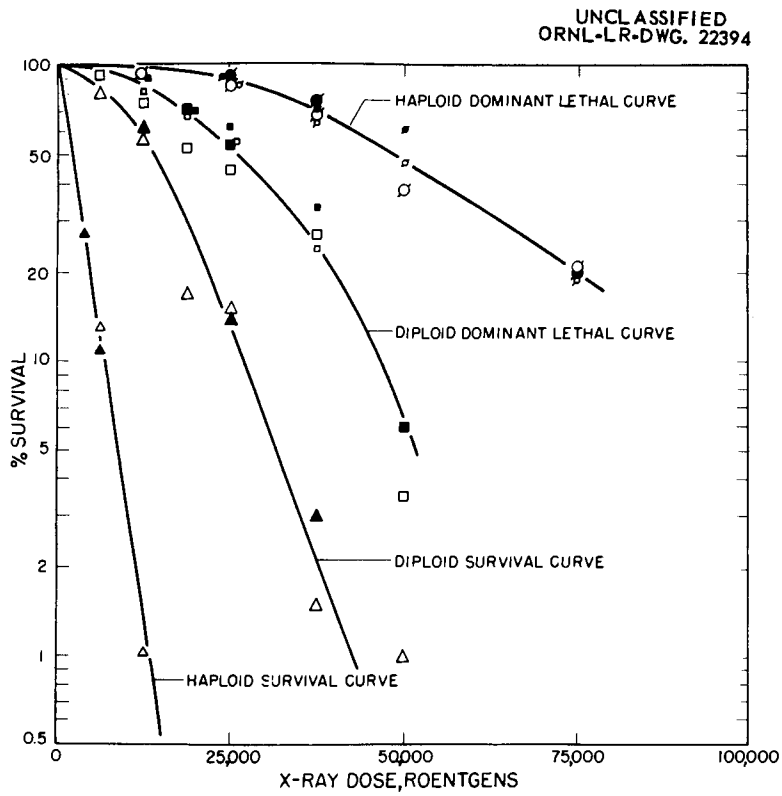

Fig. II Percentage Survival for Yeast (Owen & Mortimer, 1956)

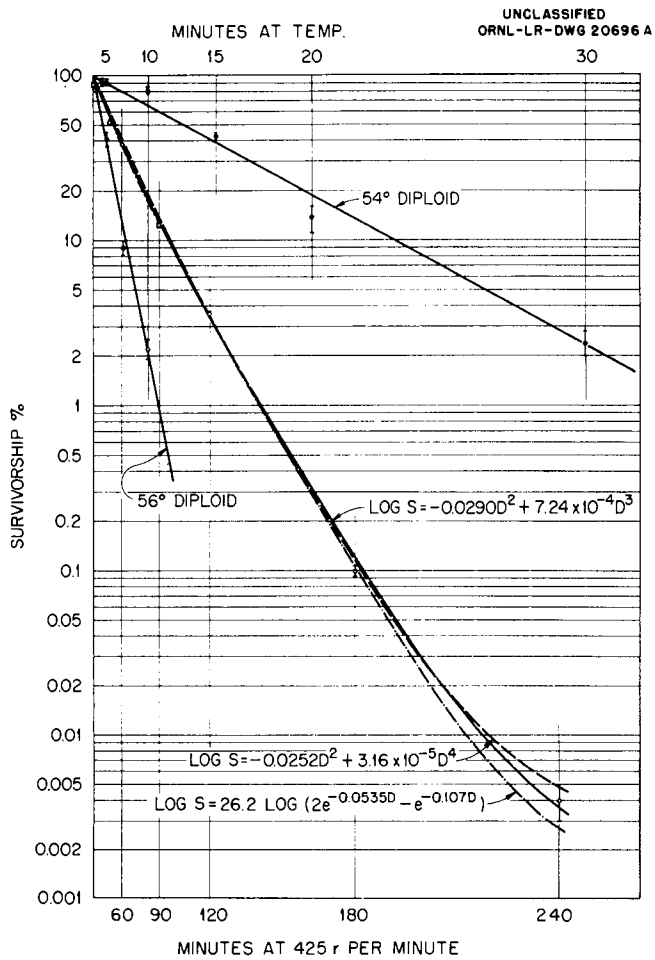

Fig. III Thermal and Radiation Killing of Yeast (Wood, 1954, 1956)

The left-hand side is the rate-of-change with λ of the survival probability. That is, $\frac{dS}{d\lambda}$ represents the rate-of-change of survival, while $\frac{1}{S}\frac{dS}{d\lambda}$ represents the probability of dying per unit of λ.

The right-hand term may be explained with the aid of Figure IV.

In a time increment $d\lambda$ the curve moves leftward, sweeping an incremental area $\rho\,(H_d,\lambda)\,dH$ from the ensemble. Hence, $\rho\,(H_d,\lambda)\,\frac{dH}{d\lambda}\Big|_{H_d}$ represents the rate at which area crosses H_d.

Eq. 20 says that the rate of change of survival probability is equal to the rate at which area beneath the ρ vs. H curve crosses from one domain to the other.

UNCLASSIFIED
ORNL LR DWG. 29394

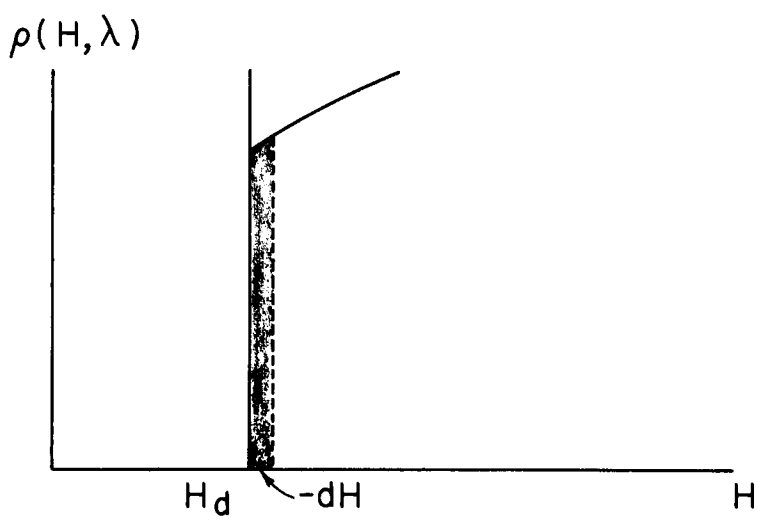

Fig. IV Increment of Area About to Cross H_d

Since $\frac{1}{S}\frac{dS}{d\lambda} = \frac{d}{d\lambda}\ln S$

$$\ln S = \int_{\lambda=0}^{\lambda=\lambda} \rho\,(H_d,\lambda)\,\frac{dH}{d\lambda}\bigg|_{H_d} d\lambda$$

The important point, however, is that there is a genetic entropy distribution, and that it shifts leftward, crossing a boundary line H_d.

The leftward movement of the curve is actually the leftward movement of each point; it represents the accumulation of read-off error in a member of the ensemble, and the consequent movement towards lower values of net entropy. As long as the organism remains to the right of H_d, it is living. When the organism crosses H_d, its net entropy falls below that required to exist in its environment, and it dies.

Thinking again in terms of the entropy distribution curve, rather than the individual points, it is clear that the distribution traces out a survival curve, or mortality pattern, as it crosses H_d. In fact, what we measure directly in a mortality experiment is the $\rho\,(H_d,\lambda)$.

The shape of the survival curve is then dependent on $\rho\,(H_d,\lambda)$ which itself depends on all intermediate $\rho\,(H,\lambda)$ back to $\rho\,(H,0)$, the shape of the entropy distribution curve at "birth."

As soon as an ensemble is born, entropy distribution commences its leftward shift, for aging produces equivocation immediately. This shift may be hastened by irradiation of the ensemble. However, since aging and irradiation produce the same type of first order chemical rate reaction in the nucleotides, the ensemble will simply age on a faster time scale. The entropy distribution, proceeding towards H_d, will assume the same intermediate configurations under radiation as under aging, but it will move faster under the influence of radiation.

Consider two identical infinite (very large) ensembles born simultaneously. If the first ensemble is given an acute radiation dose, it will move ahead of the second ensemble in its leftward shift. If it were possible, however, to "freeze" the aging of the first ensemble while the second aged normally, the second ensemble would eventually assume the same entropy distribution as the first. It will have achieved by aging what was given to the first by irradiation. If the ensembles are then allowed to age normally, they will trace out exactly the same mortality pattern in crossing H_d.

Of course, we cannot actually freeze the aging of the first ensemble, and so the second will never overtake it. However, the second ensemble differs from the first only in a time displacement; it exactly duplicates the behavior of the first - at some later time. Hence, the mortality patterns of the ensembles will differ only in a time displacement.

This type of reasoning can be applied to the description of a mortality experiment. Consider an experimenter who is attempting to duplicate a mortality experiment performed on a large number of pure-bred mice. He expects that the mortality curves will be identical.

This experiment corresponds to the following behavior of the entropy distributions:

An ensemble is born and traces out a mortality pattern. An identical ensemble is born at a later date. Since it suffers the same aging effects as the first, it traces out a mortality pattern identical with that of the first in every respect save mean time of death. The second mortality pattern will have the same dispersion, the same skewness, the same flatness, but its mean time of death will differ from that of the first by the difference in their birthdates.

The points on the first mortality curve, which represent individual organisms, may be called, starting from the left, a_1, b_1, c_1, . . . and for the second mortality curve, starting from the left, a_2, b_2, c_2, . . . If the ensembles are identical, then a_1 is interchangeable with a_2, etc.

If the first entropy distribution curve crosses H_d in the order a_1, b_1, c_1, . . . , then the second entropy distribution curve will cross in the order a_2, b_2, c_2, . . . The behavior of the organisms of the first ensemble determines the behavior of their counterparts in the second ensemble.

Therefore, each point in the mortality pattern of the first ensemble has a counterpart on the mortality pattern of the second ensemble. Since corresponding organisms die in the same order, their points in the mortality pattern represent the same percentage mortality, but different times of death.

We can therefore say that organism a_1 has suffered a certain life shortening relative to a_2, that b_1 has suffered a certain life shortening relative to b_2, etc. But organisms a_1, b_1, c_1, . . . have suffered the same life shortening, since the mortality curves of the two ensembles do not differ in dispersion, skewness, or flatness. Hence, every organism in the first ensemble has suffered essentially the same life shortening as his fellow members.

Jones (1956) has formulated a phenomonological theory of aging based on an analysis of the Gompertz function. He points out that 1 roentgen diminishes the human life span by 15 days.

There is nothing in Jones' theory to indicate whether this is an average of 15 days/r. or exactly 15 days/r. for every member of the irradiated population. Yockey's theory, however, shows that this must be the life shortening for every member, rather than an average over the irradiated members.

Thus, Yockey's theory supports Jones' work and strengthens Jones' conclusions.

IV. ANALYSIS: PROCEDURE AND RESULTS

Prefatory Remarks. It is very encouraging to discover regularities in mortality curves, such as the uniform life shortening of 15 days/r. pointed out by Jones. Regularities are often unnoticed until predicted by theory. We will later examine the regular behavior of the skewness and flatness of mortality patterns and discover a previously unnoticed regularity. It is especially encouraging that this regularity is predicted by both Jones' phenomonological theory and Yockey's information theory.

We will now describe a statistical test which verifies the prediction that acute radiation causes a uniform life shortening.

An Hypothesis. We have shown that acute irradiation should produce uniform life shortening in every member of an ensemble, so that the ensemble's mortality patterns, which is an asymmetrical (in general) bell-shaped curve, is displaced in time without distortion.

This displaced, undistorted mortality pattern must have the same dispersion, flatness, and skewness (within experimental fluctuations) as the control mortality pattern. The calculated second, third, and fourth moments about the mean, and their functions σ, β_1, β_2 must be identical (within limits of error). See Appendix I for definitions of the moments and their functions.

An Ideal Experiment. The best experiment for testing our hypothesis would have these characteristics:

1. Its members must be derived from the same strain.
2. Its members must be of the same age.
3. Its members must contain an unbiased sample of the strain at each dose level.

4. Males and females must be studied separately.

5. Its members must suffer exactly the same dose at each nominal dose level.

Numbers (1) and (4) are required to guarantee genetic similarity. Number (2) is required because this is an aging experiment. Number (3) is required to assure that the lower end of the entropy distribution (consisting of poorly endowed organisms) is not discarded. Number (5) is required because a dispersion of the doses at one nominal level will produce a spurious dispersion, or "smearing" of the mortality probability pattern.

One experiment which satisfies requirements (1) and (4) and approximates (3) will be analyzed in this paper. Operation Greenhouse, performed at the Pacific Proving Grounds in the spring of 1951, was a study of the late effects in mice of a nuclear detonation.

Operation Greenhouse. In Operations Greenhouse's gamma ray experiments (Cronkite, et. al., 1955) mice were stationed at various distances from the hypocenter of a detonation and shielded to receive a high flux gamma radiation with a very small component of fast and slow neutrons.

In the neutron experiments (Carter, Cronkite, and Bond, 1953) mice were stationed at various distances from the hypocenter and shielded to receive a high flux of above thermal neutrons with a very small component of gamma rays.

After early mortality had been recorded, these mice were transported to Oak Ridge with a set of control mice which received almost identical handling. At Oak Ridge another set of mice received x-ray exposures.

The mice were placed in individual cages, randomized in one large room where they were kept until death. Sick animals were killed in extremis.

Furth, et. al. (1954) remark:

> A significant shortening of life span resulted even from doses well beneath the threshold for acute lethality. This reduction of longevity was caused mainly by degenerative and neoplastic diseases induced or exaggerated by radiation. . . . . The reduction of longevity in proportion to dose and the observed hastening of the development of many neoplasms by irradiation suggest that ionizing irradiation accelerates the aging process. Neither aging nor radiation damage is well understood; however, the study of both and their possible interpretation is worth pursuing. Noteworthy in this connection is the observed hastening by irradiation of hereditary iris atrophy, cataract formation, vascular and interstitial sclerosis and their late sequelae, and nephrosclerosis. It is difficult to dissociate the aging process from that of injury, and most alterations attributed to aging may be those of accumulated injury. Neoplasms frequently arise in aging or injured organs. Isolated cells grown in tissue cultures do not appear to age.

Over 4500 mice of the LAF_1 strain were raised at the test site and acclimatized. At the time of exposure the mice were six to eight weeks old and sexually mature.

Mice were assigned to groups on the basis of age (which affects radiosensitivity) but not weight (which affects radiosensitivity slightly). That is, after the sexes were separated, mice of a given age group were allowed to mix freely in a large container. Then the proper number was selected at random so that each dose level would contain the same proportion of each age. All other factors thought to affect radiosensitivity were evaluated and controlled as far as possible.

Mice were irradiated simultaneously in 28 cylindrical containers which protected them from the blast, thermal radiation, and radioactive dust. Physical and biological dosimeters (film packs, Sievert ionization chambers,

Tradescantia, and spleens and thymus of other mice) were used. Tissue equivalent ionization chambers were calibrated to permit direct determination of tissue rep dose.

The individual times of death at each dose level are sufficient data for a test of our hypothesis. In tabulating the gamma ray data, the experimenters lumped together the mice at the first eight stations (920r to 711r; weighted average 749r), the second three stations (687r to 556r; weighted average 622r), the third four stations (491r to 318r; weighted average 396r) and the last two stations (287r, 192r; weighted average 240r). This lumping undoubtedly smeared the mortality pattern. Nevertheless, the Greenhouse data provides the first test of our hypothesis and we will see that despite its few less than optimum features it furnishes strong evidence in favor of the theory.

Explanation of Tables. In the last section we showed that our hypothesis is equivalent to the requirement of the constancy (within experimental fluctuations) of σ, β_1, β_2. We described the Greenhouse Experiment, and pointed out that in many respects it is a good approximation to our ideal experiment.

In Table 1 are tabulated the $\bar{x}$, σ, β_1, B_2 and their standard errors for the gamma experiments, which contain 200 to 300 members. In Table 2 and 3 are tabulated the $\bar{x}$ and σ and their standard deviations for the neutron experiments and laboratory x-ray experiments. The relatively small number of members at each dose level in the neutron and x-ray experiment invalidates (we feel) any tests of hypotheses on β_1 and β_2.

Consistency of Values. Let V_1, with standard error s_1, and V_2, with standard error s_2, be the calculated values for different mortality patterns of either σ, β_1, or β_2. Then, by hypothesis, the difference $V_1 - V_2$ is due

to experimental fluctuations. According to theory the standard deviation of the difference, s_{1-2}, is equal to $\sqrt{S_1^2 + S_2^2}$ (Worthing and Geffner, 1943, p. 198). Hence non-zero values of the parameter $Z = \frac{V_1 - V_2}{\sqrt{S_1^2 + S_2^2}}$ are, by hypothesis, due to statistical fluctuations.

If we believe that the differences $V_1 - V_2$ fluctuate normally about zero, then the probability that the two values are consistent, p(Z), may be read from a table of the tails of the normal distribution (Yule and Kendell, 1948, p. 533).

We are not certain that the differences in σ, β_1 and β_2 are normally distributed, but we will paraphrase the statistician who said that everyone believes in the normal law of error: The experimentalists think it is theoretically valid; the theorists think it is an experimental fact!

Actually, it is almost universally accepted that whether or not errors are believed normally distributed a Z value exceeding 2 or 3 is an indication of inconsistency.

Evaluation of Results. Tables 4, 5 and 6 show Z and p(Z) for the gamma ray data, and the analysis indicates consistency of σ, β_1, and β_2 in most cases. Although in some cases Z exceeds 2.0 or even 3.0, in others it is smaller than 1.0. Moreover, the extremely poor values and the extremely good ones seem to follow no pattern. We believe that the gamma ray data would appear even more strikingly consistent had the mortality patterns not been smeared when the seventeen stations were lumped together to form four nominal dose levels.

Tables comparable to 4, 5 and 6 have not been prepared for the neutron and x-ray data, since inspection of tables 2 and 3 indicates that the data are strikingly consistent.

We feel that acute irradiation with gamma rays, neutrons, or x-rays does not alter the shape of the mortality pattern.

UNCLASSIFIED
ORNL LR DWG. 29620

TABLE 1. $\bar{x}$, σ, β_1, β_2 GREENHOUSE GAMMA RAYS

Dose	N	Mean ± s. e.	σ ± s. e.	β_1 ± s. e.	β_2 ± s. e.
Female					
Control	307	123.3388 ± 1.4454	25.30 ± 1.022	0.6436 ± 0.1800	3.162 ± .3251
240r	210	106.6430 ± 1.8689	27.06 ± 1.321	0.1522 ± .12225	3.037 ± .4296
396r	290	100.9655 ± 1.8440	31.37 ± 1.304	0.3134 ± .10404	1.915 ± .1469
622r	293	87.2355 ± 1.4564	24.91 ± 1.030	0.1892 ± .12279	2.057 ± .1233
749r	266	73.1580 ± 1.3213	21.53 ± .9334	0.7269 ± .1922	2.986 ± .3081
Male					
Control	302	126.3079 ± 1.5431	26.80 ± 1.094	0.7235 ± .2533	3.592 ± .5758
240r	210	121.0475 ± 1.9249	27.87 ± 1.3608	0.2066 ± .11050	2.644 ± .2569
396r	316	112.990 ± 1.7510	31.10 ± 1.2376	0.3906 ± .26119	3.819 ± .9932
622r	302	101.1424 ± 1.4813	25.72 ± 1.0478	0.3392 ± .1779	3.327 ± .5667
749r	264	86.667 ± 1.3927	22.61 ± .985	0.4212 ± 0.03634	2.140 ± 1.1396

UNCLASSIFIED
ORNL LR DWG. 29621

TABLE 2. $\bar{x}$ and σ - GREENHOUSE NEUTRONS - EASY AND GEORGE SHOTS

Dose	Test	N	Mean ± s. e.	σ ± s. e.
Female				
Control		307	123.339 ± 1.50	25.30 ± 1.02
40 rep	G	30	107.567 ± 4.59	25.13 ± 3.25
70 rep	G	29	115.586 ± 5.18	27.88 ± 3.66
140 rep	G	18	99.111 ± 4.179	17.72 ± 2.953
180 rep	G	26	87.462 ± 6.25	31.87 ± 4.42
450 rep	G	29	76.069 ± 6.48	34.89 ± 4.58
Control		307	123.339 ± 1.50	25.30 ± 1.02
48 rep	E	25	109.280 ± 4.46	22.32 ± 3.16
139 rep	E	24	91.375 ± 6.29	30.84 ± 4.450
269 rep	E	23	92.398 ± 5.234	25.08 ± 3.70
Male				
Control		302	126.308 ± 1.54	26.80 ± 1.09
48 rep	E	25	117.320 ± 6.334	31.67 ± 4.48
139 rep	E	24	115.333 ± 6.200	30.38 ± 4.383
269 rep	E	22	107.41 ± 5.480	24.44 ± 3.68

UNCLASSIFIED
ORNL LR DWG. 29622

TABLE 3. $\bar{x}$ AND σ - GREENHOUSE LABORATORY X-RAYS

Female Dose	Mean ± s. e.	σ ± s. e.	N
Controls	123.3388 ± 1.4454	25.30 ± 1.022	307
92 r	118.856 ± 2.449	24.98 ± 1.73	104
403 r	95.135 ± 3.092	26.59 ± 2.19	74

UNCLASSIFIED
ORNL LR DWG. 29623

TABLE 4. Z AND p(Z) FOR σ - OPERATION GREENHOUSE - GAMMA RAYS

FEMALE	CONTROLS	240r	396r	622r
Controls				
240r	Z = 1.05			
	p = .315			
396r	Z = 3.16	Z = 2.33		
	p = .00158	p = 0.023		
622r	Z = 0.371	Z = 1.09	Z = 3.36	
	p = 0.72	P = 0.27	p = 0.00067	
749r	Z = 2.60	Z = 3.43	Z = 2.32	Z = 2.95
	p = 0.0093	p = 0.00057	p = 0.021	p = 0.0034

MALE				
Controls				
240r	Z = 0.632			
	p = 0.53			
396r	Z = 2.59	Z = 1.75		
	p = 0.0093	p = 0.08		
622r	Z = 0.609	Z = 1.25	Z = 3.32	
	p = 0.55	p = 0.21	p = 0.00097	
749r	Z = 2.90	Z = 3.13	Z = 5.37	Z = 2.18
	p = 0.0037	p = 0.0017	$p \rightarrow 0$	p = 0.028

UNCLASSIFIED
ORNL LR DWG. 29624

TABLE 5. Z AND p(Z) FOR β_1 - OPERATION GREENHOUSE - GAMMA RAYS

Female	Controls	240r	396r	622r
Controls				
240r	Z = 2.23 p = 0.025			
396r	Z = 1.61 p = 0.110	Z = 1.00 p = 0.317		
622r	Z = 0.747 p = 0.45	Z = 6.82 p → 0.0	Z = 0.826 p = 0.42	
749r	Z = 0.327 p = 0.74	Z = 0.071 p = 0.94	Z = 3.15 p = 0.0016	Z = 0.768 p = 0.45

Male				
Controls				
240r	Z = 0.641 p = 0.53			
396r	Z = 1.33 p = 0.183	Z = 0.650 p = 0.52		
622r	Z = 0.929 p = 0.37	Z = 0.633 p = 0.53	Z = 0.514 p = 0.61	
749r	Z = 2.72 p = 0.0069	Z = 1.49 p = 0.13	Z = 6.46 p → 0.0	Z = 1.65 p = 0.099

UNCLASSIFIED
ORNL LR DWG. 29625

TABLE 6. Z AND p(Z) FOR β_2 - OPERATION GREENHOUSE - GAMMA RAYS

Female	Controls	240r	396r	622r
Controls				
240r	Z = 0.231 p = 0.82			
396r	Z = 3.51 p = 0.00048	Z = 1.83 p = 0.072		
622r	Z = 3.35 p = 0.00097	Z = 2.18 p = 0.026	Z = 0.275 p = 0.80	
749r	Z = 0.391 p = 0.69	Z = 0.095 p = 0.92	Z = 3.10 p = 0.0019	Z = 2.78 p = 0.0055

Male				
Controls				
240r	Z = 1.64 p = 0.10			
396r	Z = 0.200 p = 0.841	Z = 1.15 p = 0.25		
622r	Z = 0.329 p = 0.74	Z = 1.11 p = 0.27	Z = 0.432 p = 0.66	
749r	Z = 2.47 p = 0.014	Z = 1.78 p = 0.072	Z = 1.68 p = 0.089	Z = 2.08 p = 0.036

V. DISCUSSION

It is remarkable that radiation should produce the samelife shortening in all members of an ensemble. Perhaps the most remarkable aspect is that the first animals to die suffer hemorrhages or gastro-intestinal disturbances, while the last animals die of neoplastic disturbances. It is almost as amazing that the mortality curves for normal aging in the control group should parallel the mortality curves for acute radiation killing at the LD 50 level (30 days). Yet these effects are demonstrated clearly by the Operation Greenhouse data.

Yockey's theory predicts certain regularities in the chronic irradiation mortality pattern. Here an organism is subjected to constant low-level irradiation until its death.

The chronic irradiated ensemble experiences an accelerated aging time scale from birth to death. Thus, it not only out distances a control ensemble in the leftward shift, but also passes across H_d more rapidly than the control ensemble. Hence, the mortality pattern of the chronic irradiated ensemble will be narrower than that of the control ensemble. Each of its moments will be smaller than that of the control ensemble. The measure of dispersion, σ, will be reduced, for it is a function of the second moment only. However, the measures of skewness and flatness, β_1 and β_2, will remain the same, for they are functions of either the third or fourth moments and the second moment, defined in such a way that they are independent of a change in time scale.

Neary, Munson and Mole (1957) have published chronic irradiation data for the mouse. We hope to undertake an analysis of the mean, dispersion, skewness and flatness of this data sometime in the near future.

APPENDIX I

Moments and Their Functions

The k-th moment about the mean of grouped data is

$$m_k = \frac{1}{n} \sum_{i=1}^{h} x_i^k f_i$$

where x_i is the deviation from the mean of the grouped data to the class mark of the i-th interval, h is the number of intervals, f_i is the frequency for the i-th interval, and $n = \sum f_i$.

Calculations are often simplified by taking moments about a convenient class mark x_0 (the working axis) and later converting to moments about the mean. Moreover, to reduce the magnitude of the numbers entering into calculations the new variable

$$u_i = \frac{x_i - x_0}{c}$$

is used, where x_i is a class mark, x_0 is the working axis, and c is the class interval.

Then the i-th moment about the mean is

$$m_k = c^k \left[\sum \frac{u_i^k f_i}{n} - k\bar{u} \sum \frac{u_i^{k-1} f_i}{n} + \frac{k(k-1)}{2} \bar{u}^2 \sum \frac{u_i^{k-2} f_i}{n} - - - \right.$$

$$\left. + (-1)^k \bar{u}^k \right]$$

where $\bar{u} = \frac{\sum u_i}{n}$ (Hoel, 1947, p. 17).

Yule and Kendall (1948, p. 161) advise the use of Sheppard's correction for grouping:

$$m_2 \text{ (corr)} = m_2 - \frac{c^4}{12}$$

$$m_3 \text{ (corr)} = m_3$$

$$m_4 \text{ (corr)} = m_4 - \frac{1}{2} c^4 m_2 + \frac{7}{240} c^8$$

A measure of central tendency is the <u>mean</u>

$$\bar{x} = x_0 + c\bar{u} .$$

A measure of dispersion about the mean is σ, the <u>standard deviation</u>

$$\sigma = \sqrt{m_2}$$

A measure of skewness, the asymmetry, is

$$\beta_1 = \frac{m_3^2}{m_2^3}$$

A measure of kurtosis, the flatness of a distribution, is

$$\beta_2 = \frac{m_4}{m_2^2}$$

In general, m_3 and m_4 are dependent on the dispersion of the distribution, but β_1 and β_2 are independent of the dispersion. For example, the Gaussian (normal) distribution has $\beta_1 = 0$, $\beta_2 = 3$ regardless of its mean or standard deviation.

We will study the functions σ, β_1, and β_2 for two reasons. First, the probable errors of the functions have been tabulated by Pearson (1930). Second, β_1 and β_2 are independent of changes in dispersion caused by changes in time scale.

Tables 7, 8 and 9 illustrate the calculation of $\bar{x}$, σ, β_1, and β_2, as well as the calculation of their probable errors.

UNCLASSIFIED
ORNL LR DWG. 29626

TABLE 7. SPECIMEN CALCULATION OF $\bar{x}$, σ, β_1, AND β_2

Class Mark	Class Freq.	u	fu	fu^2	fu^3	fu^4
25	2	-17	-34	578	-9,826	167,042
30	0	-16	0	0	0	0
35	1	-15	-15	225	-3,375	50,625
40	0	-14	0	0	0	0
45	1	-13	-13	169	-2,197	28,561
50	0	-12	0	0	0	0
55	1	-11	-11	121	-1,331	14,641
60	2	-10	-20	200	-2,000	20,000
65	2	-9	-18	162	-1,458	13,122
70	2	-8	-16	128	-1,024	8,192
75	5	-7	-35	245	-1,715	12,005
80	1	-6	-6	36	-216	1,296
85	9	-5	-45	225	-1,125	5,625
90	13	-4	-52	208	-832	3,328
95	6	-3	-18	54	-162	486
100	17	-2	-34	68	-136	272
105	22	-1	-22	22	-22	22
110	13	0	0	0	0	0
115	23	+1	23	23	23	23
120	16	+2	32	64	128	256
125	21	+3	63	189	567	1,701
130	24	+4	76	384	1,536	6,144
135	22	+5	110	550	2,750	13,750
140	27	+6	162	972	5,832	34,992
145	21	+7	147	1,029	7,203	50,421
150	20	+8	160	1,280	10,240	81,920
155	16	+9	144	1,296	11,664	104,976
160	5	+10	50	500	5,000	50,000
165	10	+11	110	1,210	13,310	14,641
170	4	+12	48	576	6,912	62,208
175	0	+13	0	0	0	0
180	1	+14	14	196	2,744	38,416
			820	10,683	41,943	774,810

UNCLASSIFIED
ORNL LR DWG. 29627

TABLE 8. SPECIMEN CALCULATION OF $\bar{x}$, σ, β_1 AND β_2
(Continued)

$\Sigma fu = 820$	$\bar{u} = 2.6678$	$x_o = 310$
$\Sigma fu^2 = 10{,}683$	$\bar{u}^2 = 7.1172$	$c = 5$
$\Sigma fu^3 = 41{,}943$	$\bar{u}^3 = 18.9873$	$n = 307$
$\Sigma fu^4 = 774{,}810$	$\bar{u}^4 = 50.6543$	

$$\bar{u} = \frac{819}{307} = 2.667752$$

$$m_2 = \frac{1}{n} c^2 \left[\Sigma fu^2 - 2\bar{u}\, \Sigma fu + n\bar{u}^2 \right] = \frac{25}{307} \left[10{,}683 - 2(2.6678)(819) + (307)(7.1172) \right]$$

$$= 692.0295 \text{ (uncorrected)}$$

$$m_3 = \frac{1}{n} c^3 \left[\Sigma fu^3 - 3\bar{u}\, \Sigma fu^2 + 3\bar{u}^2\, \Sigma fu - n\bar{u}^3 \right] =$$

$$\frac{125}{307} \left[41{,}993 - 3(2.6678)\,(10{,}683) + 3(7.1172)(819) - (18.9873)(307) \right] =$$

$$- 12{,}988.375 \text{ (uncorrected)}$$

$$m_4 = \frac{1}{n} c^4 \left[\Sigma fu^4 - 4\bar{u}\, \Sigma fu^3 + 6\bar{u}^2\, \Sigma fu^2 - 4\bar{u}^3\, \Sigma fu + n\,\bar{u}^4 \right] =$$

$$\frac{625}{307} \left[774{,}810 - 4(2.6678)(41{,}943) + 6(7.1172)(10{,}683) - 4(18{,}9873)(819) + (307) \right.$$

$$\left. (50.6543) \right] = 1{,}499{,}948.125$$

<u>Sheppard's Correction</u>

$$m_2 \text{ (corr.)} = m_2 - \frac{c^4}{12} = 692.0295 - 52.0833 = 639.9462$$

$$m_3 \text{ (corr.)} = m_3 = -12{,}988.375 = -12{,}988.375$$

$$m_4 \text{ (corr.)} = m_4 - \frac{1}{2} c^4 m_2 + \frac{7}{240} c^8 = 1{,}499{,}948.125 - 228{,}045.287 =$$

$$1{,}271{,}902.837$$

UNCLASSIFIED
ORNL LR DWG. 29628

TABLE 9. SPECIMEN CALCULATION OF $\bar{x}$, σ, β_1 AND β_2
(concluded)

$$\text{Mean} = 110 + 5(2.667752) = 123.33876$$

$$\sigma = \sqrt{m_2} = \sqrt{639.946} = 25.30$$

$$\beta_1 = \frac{m_3^2}{m_2^3} = \frac{1.687 \times 10^8}{2.621 \times 10^8} = 0.6436$$

$$\beta_2 = \frac{m_4}{m_2^2} = \frac{1.295 \times 10^6}{4.095 \times 10^5} = 3.162$$

Referring to Pearson s Tables, pages 12 - 13, and taking into account

$$n = 307, \ \sigma = 25.30$$

$$\text{p. e. mean} = (0.03850)(25.30) = 0.77405$$

$$\text{p. e. } \sigma = (0.02722)(25.30) = 0.68867$$

or

$$\text{s. e. mean} = \frac{0.97405}{.674} = 1.4452$$

$$\text{s. e. } \sigma = \frac{0.68867}{.674} = 1.0218$$

Referring to Pearson's Tables, pages 12-13, 68-71, and taking into account

$$n = 307, \ \beta_1 = 0.6936, \ \beta_2 = 3.162$$

$$\text{p. e. } \beta_1 = (0.03850)(3.15) = 0.121275$$

$$\text{p. e. } \beta_2 = (0.03850)(5.69) = 0.219065$$

or

$$\text{s. e. } \beta_1 = \frac{0.121275}{.674} = 0.1800$$

$$\text{s. e. } \beta_2 = \frac{0.219065}{.674} = 0.3250$$

APPENDIX II

A Derived Entropy Distribution

Yockey's theory does not predict the entropy distribution in an isogenic ensemble. We propose that the distribution may be calculated by information theoretical means via the calculus of variations.

Consider an infinite ensemble or organisms with a continuous distribution of information entropy. Because of fortituous gamete combinations (in sexual reproduction) or uneven division of information in fission (in asexual reproduction) daughters may be produced that have a higher information entropy than their parents.

Equivocation accumulating in the genome of parents will reduce the information transmitted to daughters. However, natural selection will tend to remove poorly endowed daughters.

Thus, although the entropy distribution is unstable and may change shape from generation to generation, the average information entropy is constrained to be constant.

Although we cannot describe the process, we believe that a stable distribution may be reached such that individual entropy variations do not affect the distribution.

Inbreeding will prevent a pure-bred strain from achieving stability. This is accomplished by selective breeding, the successive application of yes-no decisions. This is clearly the expenditure of information. (We distinguish between the bound genome entropy, which is contained in DNA, and the free inbreeding entropy, which exists in the mind of the inbreeder.) The inbreeder, if he knew the correct steps (if he had enough information) might produce a pure strain from the wild strain. Obviously he could more easily restore a slightly mongrelized strain to purity, for he would expand less free entropy.

This suggests that the stable distribution of bound entropy is the one which maximizes the free entropy. The problem is then to find p(x) to maximize

$$H(\text{free}) = \int p(x) \log_2 p(x)\, dx$$

This problem in calculus of variations arises in other information theory problems, and its solution (Goldman, 1953, p. 129) is

$$p(x) = \frac{1}{A} \operatorname{Exp}\left[-Ax\right]$$

which maximized free entropy subject to the constraint that A, the average bound entropy, is constant.

UNCLASSIFIED
ORNL LR DWG. 29629

APPENDIX III

Explicit λ Dependence of H and dH/dλ

We will now integrate Eq. 23 and substitute the result in Eqs. 17 and 20.

We have derived

$$\frac{dp_i(j)}{d\lambda} = J(\lambda)\left[-p_i(j) + \frac{1}{4}\right] \qquad \text{(Eq. 23)}$$

Integrating,

$$\int_{\lim} \frac{dp_i(j)}{p_i(j)-1/4} = \int_{\lim} -J(\lambda)\, d\lambda$$

$$\ln\left[p_i(j) - \frac{1}{4}\right]\Big|_{\lim} = -\int_{\lim} J(\lambda)\, d\lambda$$

Now the integral on the right depends on the ploidy of the organisms, for

$$J(\lambda) \doteq J_0 \qquad \text{(haploid case)} \qquad \text{(Eq. 24)}$$

$$J(\lambda) \doteq J_1\lambda \qquad \text{(diploid case)} \qquad \text{(Eq. 25)}$$

The upper limit in both cases is $\lambda = \lambda$, $p_i(j) = p_i(j)$.

The lower limit in both cases is $\lambda = 0$, $p_i(j) = p_i^0(j)$.

There are sixteen different $p_i^0(j)$, the initial values of $p_i(j)$. It is convenient, however, to assume that for $i \neq j$ all $p_i^0(j) = g \lll 1$, and for $i = j$ $p_i(j) = 1 - 3g \sim 1$.

For the haploid case, $i \neq j$,

$$\ln\left[p_i(j) - \frac{1}{4}\right]\Big|_{g}^{p_i(j)} = -\int_{\lambda=0}^{\lambda=\lambda} J_0\, d\lambda$$

$$\ln \frac{p_i(j) - \frac{1}{4}}{g - \frac{1}{4}} = - J_0\lambda$$

$$p_i(j) - \frac{1}{4} = (g - \frac{1}{4})\, e^{-J_0\lambda}$$

$$p_i(j) = \frac{1}{4} + (g - \frac{1}{4})\, e^{-J_0\lambda}$$

$$p_i(j) = \frac{1}{4}\left[1 + (4g - 1)\, e^{-J_0\lambda}\right]$$

$$p_i(j) = \frac{1}{4}\left[1 - ue^{-J_0\lambda}\right] \qquad \text{(Eq. 28)}$$

$$\frac{dp_i(j)}{d\lambda} = \frac{1}{4} J_0 ue^{-J_0\lambda} \qquad \text{(Eq. 29)}$$

where $u = 1 - 4g$

For the haploid case $i = j$

$$\frac{p_i(j) - \frac{1}{4}}{(1-3g) - \frac{1}{4}} = e^{-J_0\lambda}$$

$$p_i(j) = \frac{1}{4}\left[1 + 3(1-4g)\, e^{-J_0\lambda}\right]$$

$$p_i(j) = \frac{1}{4}\left[1 + 3ue^{-J_0\lambda}\right] \qquad \text{(Eq. 30)}$$

$$\frac{dp_i(j)}{d\lambda} = \frac{3}{4} J_0 ue^{-J_0\lambda} \qquad \text{(Eq. 31)}$$

Similarly for the diploid case $i \neq j$

$$p_i(j) = \frac{1}{4}\left[1 - ue^{-\frac{1}{2} J_1\lambda^2}\right] \qquad \text{(Eq. 32)}$$

$i = j$

$$p_i(j) = \frac{1}{4}\left[1 + 3ue^{-\frac{1}{2} J_1\lambda^2}\right] \qquad \text{(Eq. 33)}$$

The calculated $p_i(j)$ may be substituted in Eq. 17 and Eq. 20 to yield, after simplification, H and $dH/d\lambda$ as explicit function of λ. The substitution is made below for the haploid case.

$$H - H_0 = \sum_{\substack{i,j \\ i \neq j}} p(i)\, p_i(j) \log_2 p_i(j) + \sum_{\substack{i,j \\ i = j}} p(i)\, p_i(j) \log_2 p_i(j) \qquad \text{(Eq. 34)}$$

$$= \frac{1}{4}(1 - ue^{-J_0\lambda}) \log_2 \frac{1}{4}(1 - ue^{-J_0\lambda}) \sum_{\substack{i,j \\ i \neq j}} p(i) +$$

$$\frac{1}{4}(1 + 3ue^{-J_0\lambda}) \log_2 \frac{1}{4}(1 - ue^{-J_0\lambda}) \sum_{\substack{i,j \\ i = j}} p(i)$$

$$H - H_0 = \frac{3}{4}(1 - ue^{-J_0\lambda}) \log_2 \frac{1}{4}(1 + 3ue^{-J_0\lambda}) + \qquad \text{(Eq. 35)}$$

$$\frac{1}{4}(1 + 3ue^{-J_0\lambda}) \log_2 \frac{1}{4}(1 + 3ue^{-J_0\lambda})$$

$$\frac{dH}{d\lambda} = \frac{1}{\ln 2}\left\{\sum_{\substack{i,j \\ i \neq j}} p(i) \frac{dp_i(j)}{d\lambda} + \sum_{\substack{i,j \\ i \neq j}} p(i) \ln p_i(j) \frac{d}{d\lambda} p_i(j) + \right. \qquad \text{(Eq. 36)}$$

$$\sum_{\substack{i,j \\ i \neq j}} \frac{dp(i)}{d\lambda} p_i(j) \ln p_i(j) + \sum_{\substack{i,j \\ i = j}} p(i) \frac{dp_i(j)}{d\lambda} +$$

$$\left. \sum_{\substack{i,j \\ i = j}} p(i) \ln p_i(j) \frac{d}{d\lambda} p_i(j) + \sum_{\substack{i,j \\ i = j}} \frac{dp(i)}{d\lambda} p_i(j) \ln p_i(j)\right\}$$

Remembering that $\Sigma \frac{dp(i)}{d\lambda} = \frac{d}{d\lambda} \Sigma\, p(i) = 0$

$$\frac{dH}{d\lambda} = \frac{1}{\ln 2}\left\{\frac{3}{4} J_0 ue^{-J_0\lambda} + \frac{3}{4} J_0 ue^{-J_0\lambda} \ln \frac{1}{4}(1 - ue^{-J_0\lambda}) + 0\right.$$

$$\left. - \frac{3}{4} J_0 ue^{-J_0\lambda} - \frac{3}{4} J_0 ue^{-J_0\lambda} \ln \frac{1}{4}(1 + 3ue^{-J_0\lambda}) + 0\right\}$$

$$\frac{dH}{d\lambda} = \frac{3}{4} J_0 ue^{-J_0\lambda} \log_2 \frac{1 - ue^{-J_0\lambda}}{1 + 3ue^{-J_0\lambda}} \qquad \text{(Eq. 37)}$$

a. <u>Summary of λ Dependence: Haploid Case</u>

$i \neq j$

$$p_i(j) = \frac{1}{4}(1 - ue^{-J_0\lambda})$$

$$\lim_{\lambda \to 0} p_i(j) = \frac{1}{4}(1 - u) = g$$

$$\lim_{\lambda \to \infty} p_i(j) = \frac{1}{4}$$

$$\frac{dp_i(j)}{d\lambda} = \frac{1}{4} J_0 u e^{-J_0\lambda}$$

$$\lim_{\lambda \to 0} \frac{dp_i(j)}{d\lambda} = \frac{1}{4} J_0 u$$

$$\lim_{\lambda \to \infty} \frac{dp_i(j)}{d\lambda} = 0$$

$i = j$

$$p_i(j) = \frac{1}{4}(1 + 3ue^{-J_0\lambda})$$

$$\lim_{\lambda \to 0} p_i(j) = \frac{1}{4}(1 + 3 - 12g) = 1 - 3g$$

$$\lim_{\lambda \to \infty} p_i(j) = \frac{1}{4}$$

$$\frac{dp_i(j)}{d\lambda} = -\frac{3}{4} J_0 u$$

$$\lim_{\lambda \to 0} \frac{dp_i(j)}{d\lambda} = -\frac{3}{4} J_0 u$$

$$\lim_{\lambda \to \infty} \frac{dp_i(j)}{d\lambda} = 0$$

$$H - H_0 = \frac{3}{4}(1 - ue^{-J_0\lambda}) \log_2 \frac{1}{4}(1 - ue^{-J_0\lambda}) +$$

$$\frac{1}{4}(1 + 3ue^{-J_0\lambda}) \log_2 \frac{1}{4}(1 + 3ue^{-J_0\lambda})$$

$$\lim_{\lambda \to \infty} (H - H_0) = -2 \text{ bits/symbol}$$

$$\frac{dH}{d\lambda} = \frac{3}{4} J_0 u e^{-J_0\lambda} \log_2 \frac{1 - ue^{-J_0\lambda}}{1 + 3ue^{-J_0\lambda}}$$

$$\lim_{\lambda \to 0} \frac{dH}{d\lambda} = \frac{3}{4} J_0 u \log_2 \frac{g}{1 - 3g}$$

$$\lim_{\lambda \to \infty} \frac{dH}{d\lambda} = 0$$

b. <u>Summary of λ Dependence: Diploid Case</u>

$i \neq j$

$$p_i(j) = \frac{1}{4}(1 - ue^{-\frac{1}{2} J_1 \lambda^2})$$

$$\lim_{\lambda \to 0} p_i(j) = g$$

$$\lim_{\lambda \to \infty} p_i(j) = \frac{1}{4}$$

$$\frac{dp_i(j)}{d\lambda} = \frac{1}{4} J_1 u \lambda \, e^{-\frac{1}{2} J_1 \lambda^2}$$

$$\lim_{\lambda \to 0} \frac{dp_i(j)}{d\lambda} = 0$$

$$\lim_{\lambda \to \infty} \frac{dp_i(j)}{d\lambda} = 0$$

$i = j$

$$p_i(j) = \frac{1}{4}\left(1 + 3ue^{-\frac{1}{2}J_1\lambda^2}\right)$$

$$\lim_{\lambda \to 0} p_i(j) = 1 - 3g$$

$$\lim_{\lambda \to \infty} p_i(j) = \frac{1}{4}$$

$$\frac{dp_i(j)}{d\lambda} = -\frac{3}{4}u J_1\lambda e^{-\frac{1}{2}J_1\lambda^2}$$

$$\lim_{\lambda \to 0} \frac{dp_i(j)}{d\lambda} = 0$$

$$\lim_{\lambda \to \infty} \frac{dp_i(j)}{d\lambda} = 0$$

$$H - H_0 = \frac{3}{4}\left(1 - ue^{-\frac{1}{2}J_1\lambda^2}\right)\log_2 \frac{1}{4}\left(1 - ue^{-\frac{1}{2}J_1\lambda^2}\right) +$$

$$\frac{1}{4}\left(1 + 3ue^{-\frac{1}{2}J_1\lambda^2}\right)\log_2 \frac{1}{4}\left(1 + 3ue^{-\frac{1}{2}J_1\lambda^2}\right)$$

$$\lim_{\lambda \to \infty} (H - H_0) = -2 \text{ bits/symbol}$$

$$\frac{dH}{d\lambda} = \frac{3}{4}J_1 u\lambda e^{-\frac{1}{2}J_1\lambda^2}\log_2 \frac{1 - ue^{-\frac{1}{2}J_1\lambda^2}}{1 + 3ue^{-\frac{1}{2}J_1\lambda^2}}$$

$$\lim_{\lambda \to 0} -\frac{dH}{d\lambda} = 0$$

$$\lim_{\lambda \to \infty} \frac{dH}{d\lambda} = 0$$

57

c. Summary of λ Dependence: General Case

It is interesting that the substitutions

$$x = e^{-J_0\lambda} \text{ and } y = e^{-\frac{1}{2}J_1\lambda^2} \qquad \text{(Eq. 38)}$$

reduce the haploid and diploid equations to the same form.

Haploid.

$$\frac{dH}{dx} = -\frac{3}{4}u \log_2 \frac{1 - ux}{1 + 3ux} \qquad \text{(Eq. 39)}$$

$$H = H_0 + \frac{3}{4}(1 - ux)\log_2(1 - ux) + \qquad \text{(Eq. 40)}$$

$$\frac{1}{4}(1 + 3ux)\log_2 \frac{1}{4}(1 + 3ux)$$

Diploid.

$$\frac{dH}{dy} = -\frac{3}{4}u \log_2 \frac{1 - uy}{1 + 3uy} \qquad \text{(Eq. 41)}$$

$$H = H_0 + \frac{3}{4}(1 - uy)\log_2(1 - uy) + \qquad \text{(Eq. 42)}$$

$$\frac{1}{4}(1 + 3uy)\log_2 \frac{1}{4}(1 + 3uy)$$

In general

$$\frac{dH}{dZ} = -\frac{3}{4}u \log_2 \frac{1 - uZ}{1 + 3uZ} \qquad \text{(Eq. 43)}$$

$$H = H_0 + \frac{3}{4}(1 - uZ)\log_2(1 - uZ) + \qquad \text{(Eq. 44)}$$

$$\frac{1}{4}(1 + 3uZ)\log_2 \frac{1}{4}(1 + 3uZ)$$

where

$$Z = e^{-\int J(\lambda)d\lambda} \qquad \text{(Eq. 45)}$$

Bibliography

Brillouin, L.: Science and Information Theory, Academic Press, New York (1956).

Carter, R. E.; Cronkite, E. P. and Bond, V. P.: The effect of neutrons on thymic and circulating lymphocytes in the mouse. Research and Development Report USNRLD-943 NM 606 015, U. S. Naval Radiological Defense Laboratory (1953).

Cronkite, E. P.; Bond, V. P.; Lee, R. H. and Chapman, W. H.: The relative biological effectiveness of atomic bomb gamma radiation in mice. Research Report Project NM 006 012.04.86, Naval Medical Research Institute (1955).

Dancoff, S. M. and Quastler, H.: The information content and error rate of living things. Essays on the Use of Information Theory in Biology (ed. H. Quastler) 263-270, University of Illinois Press, Urbana (1953).

Furth, J.; Upton, A. C.; Christenberry, K. W.; Benedict, W. H. and Moshman, J.: Some late effects in mice of ionizing radiation from an experimental nuclear detonation. Radiology 63, 563-570 (1954).

Gamow, G.; Rich, A.; Yeas, M.: The problem of information transfer from nucleic acids to proteins. Adv. in Med. and Biol. Physics 4, 23-67, Academic Press, New York (1956).

Gamow, G. and Yeas, M.: The cryptographic approach to problems of protein synthesis. Symposium on Information Theory in Biology (ed. H. P. Yockey), Pergamon Press, London (in print, 1958).

Goldman, S.: Information Theory, Prentice-Hall, Inc., New York (1953).

Hoel, P. G.: Introduction to Mathematical Statistics, John Wiley and Sons, New York (1947).

Jones, H. B.: Factors in longevity. Kaiser Found. Med. Bull. 4, 329-341 (1956).

Lansing, A. I.: A transmissible, cumulative, and reversible factor in aging. J. Geront 2, 228-239 (1947).

Lansing, A. I.: Evidence for aging as a consequence of growth cessation. Proc. N.A.S. 34. 304-310 (1948).

Neary, G. J.; Munson, R. J.; and Mole, R. H.: Chronic Radiation Hazards. Pergamon Press, New York (1957).

Owen, M. E. and Mortimer, R. K.: Dominant lethality indiced by x-rays in haploid and diploid Sacharomyces cerebisiae. Nature 177, 625-626 (1956).

Pearson, K.: Tables for Statisticians and Biometricians, Part I, Cambridge University Press (1930).

Quastler, H. (ed.): Essays on the Use of Information Theory in Biology, University of Illinois Press, Urbana (1953).

Quastler, H.: A primer on information theory. Symposium on Information Theory in Biology (ed. Yockey, H. P.), Pergamon Press, London (in press, 1958).

Shannon, C. E. and Weaver, W.: The Mathematical Theory of Communications, University of Illinois Press, Urbana (1949).

Watson, J. D. and Crick, F. H. C.: Genetical implications of the structure of deoxyribose nucleic acid. Nature 171, 964-967 (1953).

Watson, J. D. and Crick, F. H. C.: The structure of D.N.A. Cold Spring Harbor Symposia Quant. Biol. 18, 123-131 (1953).

Watson, J. D. and Crick, F. H. C.: Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. Nature 171, 737-738 (1953).

Wood, T. H.: Influence of temperature and phase state on x-ray sensitivity of yeast. Arch. Biochem and Biophysics 52, 157-174 (1954).

Wood, T. H.: Lethal effects of high and low temperature on unicellular organisms. Advances in Biological and Medical Physics IV, 119-165 (1956).

Worthing, A. G. and Geffner, J.: Treatment of Experimental Data. John Wiley and Sons, New York (1943).

Yockey, H. P.: An application of information theory to the physics of tissue damage. Rad. Res. 5, 146-155 (1956).

Yockey, H. P.: Some introductory ideas concerning the application of information theory in biology. Symposium on Information Theory in Biology (ed. H. P. Yockey), Pergamon Press, London (in press, 1958).

Yockey, H. P.: A study of aging, thermal killing, and radiation damage by information theory. Symposium on Information Theory in Biology (ed. H. P. Yockey), Pergamon Press, London (in press, 1958).

Yule, G. U. and Kendall, M. G.: An Introduction to the Theory of Statistics. Charles Griffin and Company, Ltd., London (1948).

ORNL-2533
Health and Safety
TID-4500 (13th ed., Rev.)
February 15, 1958

INTERNAL DISTRIBUTION

1. C. E. Center
2. Biology Library
3. Health Physics Library
4-5. Central Research Library
6. Reactor Experimental Engineering Library
7-26. Laboratory Records Department
27. Laboratory Records, ORNL R.C.
28. A. M. Weinberg
29. L. B. Emlet (K-25)
30. J. P. Murray (Y-12)
31. J. A. Swartout
32. E. H. Taylor
33. E. D. Shipley
34. K. Z. Morgan
35. M. L. Nelson
36. W. H. Jordan
37. C. P. Keim
38. J. H. Frye, Jr.
39. R. S. Livingston
40. S. C. Lind
41. F. L. Culler
42. A. H. Snell
43. A. Hollaender
44. M. T. Kelley
45. A. S. Householder
46. T. A. Lincoln
47. C. S. Harrill
48. C. E. Winters
49. D. W. Cardwell
50. J. A. Lane
51-60. E. E. Anderson
61. D. Phillips
62. R. A. Charpie
63. C. E. Clifford
64. M. J. Skinner
65. R. R. Dickison
66. S. I. Auerbach
67. S. K. Allison (consultant)
68. W. H. Langham (consultant)
69. J. A. Wheeler (consultant)
70. A. Wolman (consultant)
71. ORNL - Y-12 Technical Library, Document Reference Section

EXTERNAL DISTRIBUTION

72. Division of Research and Development, AEC, ORO
73-697. Given distribution as shown in TID-4500 (13th ed., Rev.) under Health and Safety category (100 copies - OTS)