b. Cox's F' test : Let $n = n_1 + n_2$ and let the scores $t_{r,n}$ $(r = 1, 2, \ldots, n)$ denote the expected values of the order statistics of a random sample of size n from an exponential distribution with mean equal to 1. It can be shown that

$$t_{r,n} = \sum_{s=0}^{r-1} \frac{1}{n-s} \qquad (r = 1, 2, \ldots, n)$$

Combine the $p = k_1 + k_2$ uncensored observations, defined in the beginning of the section and rank them. Replace the observation with rank r with the corresponding score t_{rn} $(r = 1, 2, \ldots, n)$. If two or more of the censored observations are equal, replace each one with the average of the corresponding scores t_{rn}. Let $\bar{t}_1$ denote the average of the scores assigned to the observations fron the first sample and $\bar{t}_2$ the average of the scores assigned to the observations from the second sample. Cox [2] has shown that the ratio

$$W' = \frac{\left[k_1 \bar{t}_1 + (n_1 - k_1)\, t_{p+1,n}\right] / k_1}{\left[k_2 \bar{t}_2 + (n_2 - k_2)\, t_{p+1,n}\right] / k_2} \tag{5}$$

is approximately distributed as an F with $(2k_1, 2k_2)$ degrees of freedom, when $H_o : \lambda_1 = \lambda_2$ is true. The rejection region for the hypothesis $H_o : \lambda_1 = \lambda_2$ can be determined from the F tables. A confidence interval for $\rho = \lambda_1/\lambda_2$ can be obtained using this approximate distribution of W', as follows: Multiply each of the uncensored observations from the second sample by a fixed number ρ_o and apply Cox's procedure. This will lead to a test of the hypothesis $H_o' : \lambda_1 = \rho_o \lambda_2$.

For a given significance level α, the set of all values of ρ_0 that will lead to rejection of H_0' will from a confidence interval for ρ with confidence coefficient $1-\alpha$.

Recently, Gehan and Thomas [3] have reported a Monte Carlo study comparing the powers of the F and F' tests for small sample sizes. It was found that, when the assumption that the two samples are from exponential distributions is valid, these tests have comparable operating characteristics, (see Figure 2 below). However, if the samples are from Weibull distributions, the F test is not robust and the F' test is superior. It should be noted that the F' test requires that both sets of sample observations are censored at the same time point T_0 ; the F test is not constrained with this requirement.

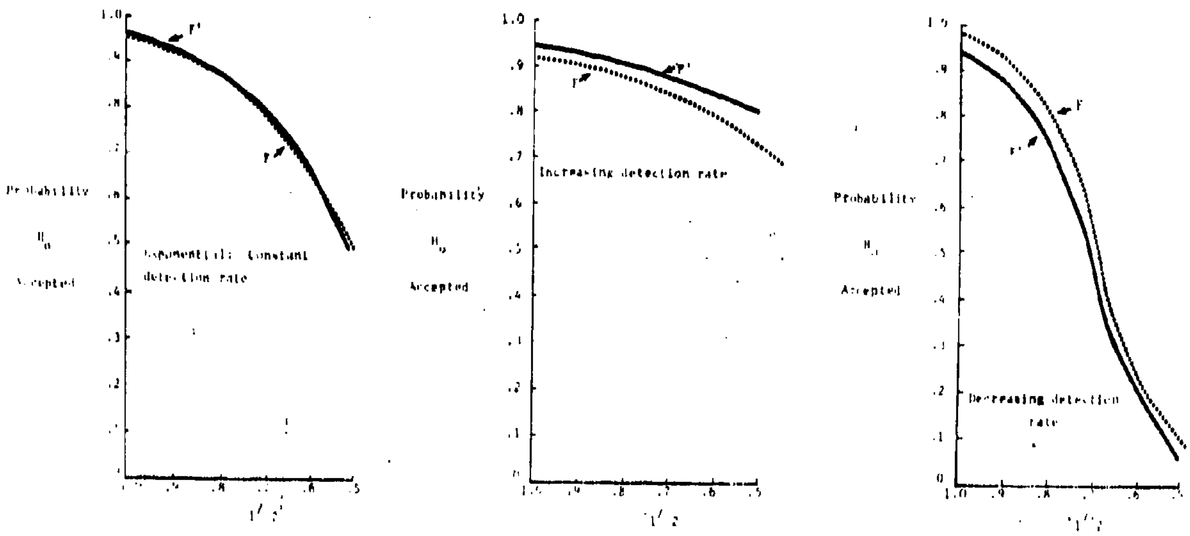


FIGURE 2. Operating Characteristic Curves for the F and F' Tests With Censored Observations for Exponential and Certain Weibull Distribution. ($n_1 = n_2 = 20$, $\alpha = .05$ and $\lambda_2 = 1.0$)

c. Confidence interval for $\psi = \lambda_1 - \lambda_2$: Suppose W_1 and W_2 are independent random variables with gamma distributions with parameters λ_1 and λ_2. Lentor and Buehler [5] obtained the conditional distribution of $H(u/v ; \psi)$ of $U = W_1$ given $V = W_1 + W_2$. This conditional distribution function involves only the parameter $\psi = \lambda_1 - \lambda_2$. A confidence interval for ψ can be obtained as follows: Set $H(u/v ; \psi)$ equal to $\alpha/2$ and $1 - \alpha/2$ respectively and solve for ψ. The two solutions for ψ will be the lower and upper confidence limits for ψ with confidence coefficient $1 - \alpha$.

The Lenter-Buehler technique can be used to derive a confidence interval for the difference $\lambda_1 - \lambda_2$ for two exponential distributions when the observations are censored. As was pointed out earlier in this paper, if the censored observations from an exponential distribution are treated as having been obtained sequentially, then $2\lambda W_k$, where W_k is the waiting time till kth arrival, given k, has a Chi-square distribution with 2k degrees of freedom. Thus W_k has a gamma distribution. We can now form two gamma distributed variables from the two sets of observations from the exponential distributions and apply Lenter-Buehler [5] technique to obtain a confidence interval for $\lambda_1 - \lambda_2$.

REFERENCES

[1] Bartholomew, D. J., The sampling distribution of an estimate arising in life testing. Technometrics 5, No. 3 pp 361-374 (1963)

[2] Cox, D. R., Some applications of exponential order scores. Journal of the Royal Statistical Society, Series B. 26, pp 103-110 (1964)

[3] Gehan, E. A., and Thomas, D. G., The performance of some two-sample tests in small samples with and without censoring. Biometrika 56, 1, pp 127-132 (1969)

[4] Halperin, M., Confidence intervals from censored samples. Annals of Mathematical Statistics. 32, 3, pp 828-37 (1961)

[5] Lentner, M. M., and Buehler, R. J., Some inferences about gamma parameters with an application to a reliability problem. Journal of the American Statistical Association 58, pp 670-77 (1963)

[6] NAVORD OD 29304/ADDENDUM. Statistical exposition of guide manual for reliability measurement program. 15 November 1967.

OPTIMIZING A FOUR-PART ASSAY PROCEDURE

Walter D. Foster
Biomathematics Division, Department of the Army
Fort Detrick, Maryland

ABSTRACT. Optimizing the procedure of a four-part assay of a bulk material is considered on the basis of the variance of each part versus its cost. In d replications (defined as Days), a settling Chamber is used c times to deposit organisms on p Plates per chamber run; the plates are incubated and Read r times per plate. A nested analysis of variance provided estimates of the components for the variance function:

$$V(\text{Assay}) = \frac{\sigma_r^2}{rpcd} + \frac{\sigma_p^2}{pcd} + \frac{\sigma_c^2}{cd} + \frac{\sigma_d^2}{d}$$

The cost function is the sum of the cost rate x number of repeats for each part:

$$\text{Cost} = (\text{unit Reading cost})\ rpcd + (\text{unit Plating cost})\ pcd$$

$$+ (\text{unit Chamber cost})\ cd + (\text{unit Day cost})\ d.$$

A Lagrangian multiplier approach yielded minimum variance for a fixed total cost:

$$\text{Min}\ [V(\text{Assay}) + \lambda(\text{Cost})].$$

This approach is contrasted with non-linear programming and integer programming.

INTRODUCTION. A problem common to both bulk sampling and sample surveys is to obtain as much information about a population, usually stratified or segmented in time or space, as possible for a given budget. For example, grain production in the mid-west or even from a specific farm would need to be sampled according to various strata defined by bags, bushels, wagon loads, etc. To maximize the information, i.e., minimize the variance of the sample mean for a given cost is the way the problem is usually stated.

Duncan[1], among others, has written extensively on procedures for bulk sampling and has a pertinent bibliography. Similarly, Cochran[2], to pick another outstanding writer, has worked extensively in the area of sample surveys. The problem addressed here is to optimize an assay procedure for a sporulating organism where the steps may be considered as a three-phase procedure. This problem is answered partly by borrowing the technique from bulk and sample survey solutions; it is also approached from the point of view of integer programming.

A "day" is a natural block of work. A small amount of the spore material in dry state is encapsulated and dispersed with a CO_2 pistol into a small chamber. Several agar plates are set in the bottom of the chamber onto which the dispersed spores fall. After a suitable incubation period in a humidity chamber, the plates are "read" until 100 organisms are counted and the percent viability is estimated. Several such readings are made. It is the purpose here to minimize the variance of this assay procedure subject to a particular assay cost.

I. VARIANCE FUNCTION. An experiment was designed specifically to provide estimates of the various sources of variation. In each of several "days," repeated chamber runs provided an opportunity to observe variation from one chamber run to another within the same day. Six chamber runs per day were scheduled, three agar plates per chamber run were prepared, and each plate was read six times. The experiment continued over 18 days. For sake of simplicity the variation due to technicians is omitted in this report.

A nested analysis of variance gives both the structure of the analysis and the results.

ANALYSIS OF VARIANCE ON % VIABILITY

Source	df	MS	EMS
Days	17	315.013	$\sigma_R^2 + 2\sigma_p^2 + 6\sigma_c^2 + 36\sigma_D^2$
Chambers in Days	90	26.990	$\sigma_R^2 + 2\sigma_p^2 + 6\sigma_c^2$
Plates in chambers	216	20.981	$\sigma_R^2 + 2\sigma_p^2$
Readings in plates	324	15.012	σ_R^2

$$\sigma_R^2 \doteq 15; \quad \sigma_p^2 \doteq 3; \quad \sigma_c^2 \doteq 1; \quad \sigma_D^2 \doteq 8$$

Because of the balanced arrangement, there was no problem in estimating the components in this hierarchical analysis according to the expected mean square shown in the table above. The estimates of these components of variance are also shown there. In this completely randomized design with all factors considered random, the variance of the mean percent viability is given the following equation:

$$V(\overline{\%V}) = \frac{\sigma_R^2}{r} + \frac{\sigma_p^2}{rp} + \frac{\sigma_c^2}{rpc} + \frac{\sigma_D^2}{rpcd}$$

This variance function permits the experimenter to obtain any degree of precision he is willing to pay for merely by assigning appropriate values to the number of readings per plate, the number of plates per chamber, the number of chamber runs per day, and the number of days. One criterion for this degree of accuracy is the cost of the assay. A second criterion concerns the intended use of the material and to what degree of accuracy he must know its viability.

II. COST FUNCTION. Cost functions are notoriously difficult to describe realistically, especially when the assay procedure itself is conducted on a sporadic basis. The function chosen here was selected primarily because of its simplicity and was felt to be adequate for the problem at hand:

$S = rpcd\,R + pcd\,P + cd\,C + d\,D$, where $R = .20$; $P = .50$; $C = .50$; and $D = 2.50$ are the (fictitious) cost per unit for Reading, Plate, Chamber and Day, and r, p, c, d are the number of units of each.

As is clear, this function also follows a nested construction and assigns to each portion of the assay a fixed cost so that total cost is obtained merely by counting the number of component parts. Overhead costs such as equipment, utilities, laboratory space, etc., have been included in the "day" costs.

If the experimenter is willing to pay S dollars for one assay of the material, then the conjunctive use of the variance and the cost function will optimize the assay procedure for that cost. Conversely, the cost function will show him what he would need to pay to obtain an assay for a prescribed precision. The optimization processes are shown in the next section.

III. OPTIMIZATION PROCESSES.

A. Lagrangian Multipliers

Let us write the following function

$$L = V(\overline{X}) + \lambda S$$

which includes both the variance and the cost as a quantity to minimize. Following the standard procedure, partial derivatives of the function L are taken with respect to the parameters d, c, p, r giving four equations in five unknowns, the fifth unknown being the variable λ. The fifth equation is the cost function itself. These derivatives, which can be seen in the Appendix, give rise to a set of five non-linear simultaneous equations whose solutions are also shown in the Appendix and were obtained as a unique set by virtue of the simplicity of the nested design.

It is clear, although not proved mathematically, that any hierarchical design with random effects and an associated nested cost function will have these two properties: non-linearity and analytic solution.

The procedure involving the Lagrangian multipliers clearly treats d, c, p, r as continuous variables. Obviously they must be positive as well as all of the other input values such as cost rates and the variance components. In keeping with this approach, we should then expect to find the solution for these variables to be non-integers. For a limit of S = \$9.00 per assay, the optimum solution gave

$$d = 1.98; \quad c = .79; \quad p = 1.73; \quad r = 3.54; \quad V(\bar{X}) = 7.4; \quad S = 9.0$$

Because the experimenter must work with integer values of these variables, the next highest and lowest integer value of each variable were examined in combinations with similar upper and lower values for all other variables giving rise in general to 2^4 combinations. The variance and cost functions were computed for 8 of the 16 combinations plus one other; and, the one that minimized the variance for the allotted cost was chosen: specifically where S = \$9.00, d = 2, c = 1, p = 1, r = 5. These 9 combinations are shown in the table below with the selected set shown with an asterisk.

d	c	p	r	$V(\bar{X})$	S
1	1	1	3	17.0	4.1
1	1	1	4	15.8	4.3
1	1	2	3	13.0	5.2
1	1	2	4	12.4	5.6
2	1	1	3	8.5	8.2
2	1	1	4	7.9	8.6
2	1	2	3	6.5	10.4
2	1	2	4	6.2	11.2
2	1	1	5	7.5	9.0*

By extending the perturbation range beyond adjacent values, such a table also has the advantage of displaying to the experimenter what a small increase in cost would yield in terms of increased precision. Similar solutions would be available at other levels of possible cost, S, per assay.

B. Integer Programming

Integer programming is that type of mathematical programming that seeks to optimize a particular function subject to certain constraints where the constraints are expressed in integer form. Mathematical theory to date has not discovered a general procedure for problems in integer programming but several alternative procedures are available that are felt to be optimum for well-behaved surfaces. A fuller discussion of integer programming is beyond the scope of this paper and is not pursued here. An effective alternative procedure consists of asking the computer to map the surface for all combinations which are reasonable and to select the optimum value which satisfies the cost restraint.

APPENDIX

(1) $S = Dd + Cdc + Pdcp + Rdcpr$

(2) $L = [V(\bar{X}) + \lambda(S)] = \frac{\sigma_d^3}{d} + \frac{\sigma_c^3}{dc} + \frac{\sigma_p^3}{dcp} + \frac{\sigma_r^2}{dcpr}$

$$- \lambda S + \lambda Dd + \lambda Cdc + \lambda Pdcp + \lambda Rdcpr$$

(3) $\lambda = \frac{\sigma_r^2}{Rd^2c^2p^2r^2}$

(4) $\frac{\partial L}{\partial p} = \frac{-\sigma_p^2}{dcp^2} + \frac{-\sigma_r^2}{dcp^2r} + \lambda[Pdc + Rdcr] = 0$

(5) $\frac{\partial L}{\partial c} = \frac{-\sigma_c^2}{dc^2} + \frac{-\sigma_p^2}{dc^2p} + \frac{-\sigma_r^2}{dc^2pr} + \lambda[Cd + Pdp + Rdpr] = 0$

(6) $\frac{\partial L}{\partial d} = \frac{-\sigma_d^2}{d^2} + \frac{-\sigma_c^2}{d^2c} + \frac{-\sigma_p^2}{d^2cp} + \frac{-\sigma_r^2}{d^2cpr} + \lambda[D + Cc + Pcp + Rcpr] = 0$

(7) $r^2 = \frac{P}{R} \cdot \frac{\sigma_r^2}{\sigma_p^2}$

(8) $p^2 = \frac{C}{P} \cdot \frac{\sigma_p^2}{\sigma_c^2}$

(9) $c^2 = \frac{D}{C} \cdot \frac{\sigma_c^2}{\sigma_d^2}$

(10) $d = \dfrac{S}{D + C\sqrt{\frac{D}{C} \cdot \frac{\sigma_c^3}{d^2}} + P\sqrt{\frac{D}{P} \cdot \frac{\sigma_p^3}{d^2}} + R\sqrt{\frac{D}{R} \cdot \frac{\sigma_r^2}{d^2}}}$

(11) $d = \dfrac{\sqrt{D} \cdot \sigma_d \cdot S}{D[\sigma_d\sqrt{D} + \sigma_c\sqrt{C} + \sigma_p\sqrt{P} + \sigma_r\sqrt{R}]}$

AN APPLICATION OF MATHEMATICAL PROGRAMMING TO EXPERIMENTAL DESIGN

J. Richard Moore
U. S. A. Aberdeen Research and Development Center
Aberdeen Proving Ground, Maryland 21005

It often happens that a number of observations (measurements) of interest can be made each time a complex system is operated. When there are several different modes in which such a system can be operated (missions) the set of observations may be different for each mission. Many observations may be common to several modes of operation while some may be peculiar to a particular mission.

When it is necessary to obtain a number of observations for each of a known set of subsystems (or functions) the question of how many times the system must perform each type mission arises. A reliability demonstration test for a complex system is an example of the above. This paper gives an example of a procedure by which the required set of observations can be obtained at a minimum cost. Hopefully this example will help those who are active in research and development to recognize situations in which appreciable saving can be effected by the use of mathematical programming. These procedures are relatively simple to use once the problem to which they apply has been identified.

As an example, consider a complex system which consists of six major subsystems. A reliability demonstration test must be conducted for the system and also for each of the subsystems operating as part of the system. This system is capable of performing four different types of mission. The amount of operating time (number of operating cycles) for each subsystem during a mission depends on the mission type but it is assumed that the probability that a specific subsystem will fail during a mission depends only on the amount of operating time (cycles) it accumulates during the mission and not on the type of mission which the system is performing. That is, it is assumed that the distribution of time (cycles) to failure does not depend on the type of mission the system is performing when the subsystem is accumulating operating time (cycles).

It is necessary to use a reliability demonstration plan which allows no failures because of the limited availability of test resources. For the system to be tested this dictates that, in order to demonstrate the desired reliability with the required confidence, each of the subsystems must accumulate the operating experience given in Table I.

TABLE I

Required Subsystem Operating Experience

Subsystem	Operating Time (Cycles)
1	60 minutes
2	30 minutes
3	40 minutes
4	80 minutes
5	20 cycles
6	30 cycles

It is further required that the system must complete at least two of each type mission without failure.

The operating time (cycles) for each subsystem during each type mission is given in Table II

TABLE II

Operating Experience for Subsystems by Mission Type

Subsystem	Mission Type 1	2	3	4
1	6	12	0	3
2	9	0	4	6
3	8	2	0	5
4	0	6	10	3
5	5	1	2	0
6	3	0	2	3

The units of measurement for the entries in Table II are minutes for subsystems one through six and cycles for subsystems seven and eight. Thus, Table II says that a subsystem one operates for six minutes during a Type 1 mission, twelve minutes during a Type 2 mission, three minutes during a Type 4 mission and doesn't operate during a Type 3 mission, etc.

The problem is to find the best way to run this test program in the sense that the total cost of testing is minimized. The cost of running each type mission is given in Table III.

TABLE III

Cost of Running Each Type Mission

Mission Type	Cost of Mission
1	40
2	38
3	43
4	39

The cost data in Table II are relative costs which are in thousands of dollars in this example.

With the preceding information available, it is seen that this is a problem in linear programming in which one must determine the minimum number of times to require the system to complete each type mission, i.e., minimize the cost of testing, subject to the constraints that each subsystem must accumulate at least as much operating experience as is given in Table I and also the system must complete at least two of each type mission. Since the solution will be in terms of number of tests, it also follows that the solution vector for the linear program must be integer valued. A problem of this type is called an integer programming problem.

The first step in finding a solution to the problem will be to formulate it in standard terminology. Let Y be a column vector with jth element equal to the number of times the jth type mission must be run, that is

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

where $\{y_j\}$ are unknown which must be determined. Let C be a row vector with elements equal to the mission costs given in Table III, B be a column vector with elements equal to the required total operating requirement for each subsystem as given in Table I and A be a matrix with elements equal to the operating time (cycles) for each mission type as given in Table II. With this notation the problem is to find the vector Y such that:

$$CY = \text{minimum}$$

$$AY \geq B$$

$$Y \geq \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix}$$

and all of the elements of Y must be integers. Substituting the numbers given in Table I, Table II and Table III in the above relationships leads to the following statement of the problem:

Find a vector of integers, Y, such that

$$CY = 48y_1 + 38y_2 + 43y_3 + 39y_4 = \text{minimum}$$

and

$$\begin{aligned} 6y_1 + 12y_2 + 3y_4 &\geq 60 \\ 9y_1 + 4y_3 + 6y_4 &\geq 30 \\ 8y_1 + 2y_2 + 5y_4 &\geq 40 \end{aligned}$$

$$A \cdot Y = \begin{aligned} 6y_2 + 10y_3 + 3y_4 &\geq 80 \\ 5y_1 + y_2 + 2y_3 &\geq 20 \\ 3y_1 + 2y_2 + 3y_4 &\geq 30 \end{aligned} = B$$

$$Y = \begin{aligned} y_1 &\geq 2 \\ y_2 &\geq 2 \\ y_3 &\geq 2 \\ y_4 &\geq 2 \end{aligned}$$

The last four constraints can be eliminated from the problem by the translation of axis

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = X = Y - \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} = Y - 2e$$

$$\text{where } e = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

This transforms the original problem to the equivalent problem: Find a vector X such that

$$CX = \text{minimum}$$

$$AX \geq B_x$$

$$X \geq 0$$

where $B_x = B - AM$

ahd $M = 2e$ for this problem.

The next step is to inspect the vector B_x to determine whether the complexity of the problem can be reduced. If any element of B_x is zero or negative the operating requirements for the corresponding subsystem will have been satisfied when the system has completed the number of each type mission necessary to satisfy the constraints $Y \geq M$. Thus, the size of the problem can be reduced by dropping all constraints corresponding to zero or negative elements of B_x.

In the example $B_x = B - 2Ae$ is

$$B_x = \begin{pmatrix} 60 \\ 30 \\ 40 \\ 80 \\ 20 \\ 30 \end{pmatrix} - \begin{pmatrix} 6 & 12 & 0 & 3 \\ 9 & 0 & 4 & 6 \\ 8 & 2 & 0 & 5 \\ 0 & 6 & 10 & 3 \\ 5 & 1 & 2 & 0 \\ 3 & 0 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 18 \\ -8 \\ 10 \\ 42 \\ 4 \\ 14 \end{pmatrix}$$

The second element in B_x is negative, which means that if each type mission is run two times in order to satisfy the requirements $X \geq 0$ ($Y \geq 2e$) there will be $(9)(2) + (0)(2) + (4)(2) + (6)(2) = 38$ minutes operating time on subsystem number two, which is eight minutes more than the required thirty minute operating time for that subsystem; hence, the second element in B_x is minus eight. This allows a reduction in the size of the problem without changing the solution by dropping the second constraint equation.

Letting A_0 and B_x° represent the reduced matrix and vector resulting from the deletions, the problem is rewritten as:

Find a vector X such that

$$CX = \text{minimum}$$
$$A_0 X \geq B^\circ_x$$
$$X \geq 0$$

The next step is to inspect the columns of the matrix A_0 to determine whether the dimension of the problem can be reduced by omitting one or more of the elements of the vector X. If any column of A_0 contains all zeros then none of the subsystems for which data is needed will operate during the type mission corresponding to that column. Requiring the system to complete such a mission would only increase the cost of testing without generating any needed data. The dimension of the problem is reduced by eliminating the zero columns from A_0 and the corresponding elements from the vectors X and C. Letting A°, X° and C° be the new matrix and vectors which result from this transformation, the problem is stated in final form as: Find a vector X° such that

$$C^\circ X^\circ = \text{minimum}$$
$$A^\circ X^\circ \geq B^\circ_x$$
$$X^\circ \geq 0$$

In the numerical example, omitting the second row in A and the second element in B_x reduces the problem to the form:

$$CX = \begin{pmatrix} 40 & 38 & 43 & 39 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \text{minimum}$$

$$A_0 X = \begin{pmatrix} 6 & 12 & 0 & 3 \\ 8 & 2 & 0 & 5 \\ 0 & 6 & 10 & 3 \\ 5 & 1 & 2 & 0 \\ 3 & 0 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \geq \begin{pmatrix} 18 \\ 10 \\ 42 \\ 4 \\ 14 \end{pmatrix} = B_X$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The dimension of the problem can not be reduced because none of the columns of A_0 contain only zero elements. Thus, it is in the final. reduced form.

With the exception of the requirement that the elements of X° must be integers, the above is simply a linear programming problem which can be solved using the simplex method. It can be shown that, if all of the elements of A° are either zero or one and all of the elements of B°_X are integers, then the solution vector obtained using the simplex method will have all integer elements. This would be the case if each of the subsystems were of cyclic nature, e.g., relays which operate for only one cycle when required during a mission. Also, if for each subsystem, the operating time (number of cycles) is the same during each mission in which it is required to operate and the operating requirement is a multiple of this, then the problem can be reduced to the zero-one-integer problem and the simplex method can be used to find the solution vector for the integer programming problem. It should be noted that the above remarks apply to the problem after it has been put in reduced form by making. all possible decreases in size and dimension. That is, even though it is not possible to formulate the original problem as a zero-one-integer problem it may be possible to do so for the problem in reduced form.

It should be emphasized that the conditions of the preceding paragraph are sufficient but not necessary to assure that the simplex method will yield an integer solution to the problem. Since the simplex method is much easier to apply than the methods of integer programming, it is advisable to first find the simplex solution and see whether it is integer valued before proceeding further.

The simplex solution for the example given here is not integer valued so a search routine was used on the computer to find the solution, which is:

$$X = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 3 \end{pmatrix} \text{ or } Y = X + 2e = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 5 \end{pmatrix}$$

with minimum cost of

$$CY = (40 \quad 38 \quad 43 \quad 39) \cdot \begin{pmatrix} 2 \\ 3 \\ 5 \\ 5 \end{pmatrix} = 604$$

The reader may find it interesting to show that the solution satisfies all of the constraints.

The problem formulated here is of the same structure as the nutrition or diet model found in the literature. The reader is referred to Chapter 27 of "Linear Programming and Extensions" by George B. Dantizg for an excellent treatment of the formulation and treatment of that problem.

Transmission of Infrasonic Waves Generated by Large Missile Launches

by

R. E. Lacy and C. E. Sharp
Institute for Exploratory Research
U. S. Army Electronics Command
Fort Monmouth, New Jersey

Summary

Infrasonic pressure waves generated during launching of large missiles can be detected at distances greater than 2000 kilometers. On a number of occasions during recent years, two infrasonic wave trains spaced about 30 minutes apart have been recorded by sensor array areas located at Fort Monmouth, N. J. These areas are approximately 1400 kilometers from the launch pads at Cape Kennedy, Florida. To date, no explanation for these observed early arrivals of an infrasonic wave train which appears to travel at approximately twice the normal velocity of sound through the atmosphere, has proven to be acceptable. Most recently, we have observed a relationship between these sounds and the jet streams traveling the proximity of the path from Cape Kennedy to Fort Monmouth. On this basis, we have evolved a hypothesis that such correlation exists between the jet streams and this anomalous propagation of sound waves. Because of the large amount of jet stream data, but relatively few missile-firing events, a statistical design of further experiments adequate to test the hypothesis is planned. A theoretical analysis explains the pressure wave spectra received as consisting of three separate groups of spectra.

Introduction

Infrasonic wave propagation phenomena have been explored with renewed interest in recent years. In this report we summarize our current findings. We are concerned here only with infrasonic frequencies traversing the atmosphere. Infrasonic waves are inaudible sound waves whose frequency of oscillation are below 15 Hz; they have the same velocity as audible sounds, i.e., about 332 m/sec at 20°C and a pressure of 760 mmHg. The absorption of infrasound is considerably less than absorption of audible sound of the atmosphere due to heat conduction and viscosity.[1] Because of this, the detection at long distances of infrasonic energy generated by the large missiles, as represented by Titan III and Saturn V, launched from Cape Kennedy, Florida, is possible. Frequency analysis of a number of magnetic tape recordings made at Fort Monmouth, N. J. during the arrival of infrasonic waves generated by the missile launches show the maximal energy propagated to be centered within 0.5 to 5 Hz. The maxima of these traveling pressure oscillations recorded in the Fort Monmouth test areas seldom exceed 2.5 dynes/cm^2. The "normal" traveltime for infrasonic waves from Cape Kennedy to Fort Monmouth is between 70 to 85 minutes after launch. This

This article has been reproduced photographically from the author's manuscript.

variation in arrival time may be due to the directions and velocities of the surface winds along the travel path. The duration of the infrasonic wave trains also has been observed to vary between five and ten minutes. This variation may be the result of the relative signal-to-noise ratio at the sensor location for each launching. Furthermore a so-called precursor or wave train has been sporadically observed that arrived about 10 minutes earlier than the "normal" wave train. Figure 1 clearly shows this precursor recorded, when the wind turbulence was practically nil, at a site 1000 km north of Cape Kennedy, Florida.

The Wind Turbulence Problem

The detection at long distances of acoustical energy produced by the large missile launches (Titan, Atlas, Saturn, etc.) from Cape Kennedy is complicated by low signal-to-noise ratios. The acoustical energy to be detected is in the form of atmospheric pressure waves of usually less than 2 dynes/cm^2, contributed in the selected frequency spectra 0.01 to 10 Hz. At the distance of interest for this study, approximately 1400 km, the amplitude and frequency range of the pressure waves detected have been decreased considerably by atmospheric absorption, by the dispersive effects of wind and temperature variations in the lower and upper atmospheric regions, and by spherical spreading. At long ranges, the combined effect of these factors results in a reduction of the amplitude of the missile-produced atmospheric pressure waves to less than the random local atmospheric pressure variations, ambient at the sensor location. These ambient variations due to local atmospheric conditions can have pressure amplitudes in the order of 100 to 500 dynes/cm^2 and frequency spectra similar to missile-produced waves.

We found the most promising approach used to overcome these difficulties, when the position of the infrasonic source is known, to be in the employment of a large number of sensors (microphones, microbargraphs, or pressure transducers) in a linear array, parallel to the infrasonic pressure wave front and the use of real-time correlation techniques. The output of each sensor in the array is combined with the outputs of the others by summation or multiplication. If the sensors are sufficiently spaced in the linear array, the pressure variations caused by wind turbulence at each sensor tend to produce uncorrelated outputs. These uncorrelated outputs can be expected to sum power-wise to NA^2, where N is the number of sensors and A is the output amplitudes; whereas the missile-produced pressure waves arriving in phase at all sensors in the broadside linear array will sum up to $(NA)^2$. The improvement in the signal-to-noise ratio over that of a single sensor will be equal to $10 \log (NA)^2/NA^2$ or in dB equal to $10 \log N$. The greater the number of sensors in the array, of course, the greater the detection improvement[2] (Fig. 2). In addition, an effective wind screen (Fig. 3) has been devised to inclose each sensor. This screening can result in a further improvement in signal-to-noise ratio of 10 to 30 dB. The amount of improvement is a function of the degree of wind turbulence.

A 1000' broadside linear array of 20 screened and equally spaced microphone sensors have been in operation for several years at a test site in Wayside, N. J. (Fig. 4). In order to verify the direction of arrival and determine the velocity of propagation, a second linear array of 10 sensors is in operation at a site in Middletown, N. J., seven miles north of Wayside. Both arrays are arranged parallel to the acoustic-pressure wave-fronts arriving from Cape Kennedy. These arrays are remotely controlled and the outputs recorded via telephone circuits terminating in a laboratory room. The time of pressure wave arrival from launching time T_0, the wave durations, velocity, and other pertinent data are recorded for each array and each scheduled large missile launching.

Observations

The occasionally observed early arrival (precursor) of infrasonic waves in the time frame of 38 to 43 minutes (Fig. 1) preceding the normal arrival period is of considerable interest to those concerned with acoustic propagation phenomena. Observed amplitudes and durations of these early arrivals are about one-half that of the normal arrival. In the lower atmospheric regions, the absorption coefficient, which is a function of temperature, viscosity, and pressure, is extremely small at these low acoustic frequencies. In the upper atmospheric regions, the absorption coefficient is greater, principally because of the lower atmospheric pressure. This lends support to our hypothesis that upper atmospheric ducting is responsible for the early arriving pressure waves at the jet stream level (30 000 to 40 000'). Apparently, early arrival occurs only when the jet stream flows northward along the east coast from the direction of Florida. We have observed this phenomena during the winter months and it has been confirmed by studies of the tropopause wind analysis stream function charts published daily by the ESSA Environmental Data Service.[3] The velocity of the jet stream as given in these charts plus the velocity of sound at the relevant altitude could account for the early arrival of the infrasonic wave train. This theory is further confirmed by comparing the observed reduced wave amplitudes and durations of the early arrivals with the amplitudes and durations of the normal arrivals. A four-minute film showing the variation of these wind streams over a period of a year will accompany this presentation. These data show that the jet streams prevail in a west to east direction across the U.S.A. in the summer. As late fall occurs, the streams drop to the southern portion of the country with an attendant curvature to the north along the east coast. Figure 5 shows typical west to east summertime patterns, while Figure 6 shows the winter streams approximately parallel to the east coast. The latter is thereby in the correct orientation to account for the precursor, or earlier than normal wave arrival.

Design of Experiment-Statistical Problem Areas

During observations of the early infrasound arrivals from Cape Kennedy, the detection of the weak infrasonic wave trains from long distances is often masked by the effects of wind turbulence, particularly during the windy winter months in the northeastern part of the U. S. The broadside array of many sensors (100 or more) appears to be the most promising solution to the reduction of the effect of wind turbulence. The microphone sensors presently employed require considerable maintenance and therefore large arrays of a hundred or more sensors become rather impractical. At the present time, the problem of further reducing the effects of wind turbulence remains with us. Investigation of two types of infrasonic gradient[4] microphone sensors are currently in progress. The gradient types offer promise of greater reliability, lower costs per unit, and may be less affected by wind turbulence.

Then the problem of site selection for the sensors further broadens the experimental design problem. With but only two or three site instrumentations economically allowed, the question arises as to where the sampling would prove to be most profitable in terms of helpful data. Additionally, the correlation of the sparse experimental data with the voluminous wind data provide problem areas which may be amenable to statistical techniques. As shown in the accompanying film and Figures 5 and 6, the wind data shows distinct summer-winter variations. But, the launching of missiles are relatively few and far between.

To further complicate the real situation, there could be and are, other hypotheses attempting to explain the relative occurrences of the normal and precursor wave trains. For this experimental design consideration, though, we are confining our attention to the one hypothesis, i.e., the jet stream hypothesis. We plan to arrange experiments to confirm, or deny, this hypothesis. Initially, therefore, we must determine how best to proceed in this endeavor that is typical of many such quests in what might be termed macroscopic experimental research involving sporadic and uncontrollable conditions.

References

1. Richard K. Cook, "Strange sounds in the atmosphere, Part 1," Sound, Vol. 1, No. 2, pp. 12-33, May-June 1962.

2. F. B. Daniels, "Noise-reducing line microphone for frequencies below 1 cps," J. Acoust. Soc. Amer., Vol. 31, No. 4, 529-531, April 1959.

3. "Tropopause wind analysis stream function charts," U. S. Department of Commerce, ESSA, Environmental Data Service, Weather Bureau, Silver Springs, Maryland 20910.

4. B. B. Bauer, "Design and measurement of infrasonic gradient microphones and wind screens," J. Acoust. Soc. Amer., Vol. 44, No. 5, pp. 1428-1436, 1968.

Selected Bibliography

1. "Launch Acoustical Environmental Summaries," Published by the John F. Kennedy Space Center, NASA, TR 927, 30 January 1969, Office of the Director of Information Systems.

2. C. T. Johnson and F. E. Hale, "Abnormal sound propagation over the southwestern United States," J. Acoust. Soc. Amer. Vol. 25, No. 4, pp. 642-650, July 1953.

3. William L. Donn, Eric Posmentier, Uri Fehr, N. K. Balachandran, "Infrasound at long range from Saturn V, 1967," Science, Vol. 162, pp. 1116-1120, 6 December 1968.

4. Harold N. Ballard and Mike Izquierdo, "Reduction of Microphone Wind Noise by the Generation of a Proper Turbulent Flow," Environmental Sciences Div. Report ERDA-262, February 1965.

APPENDIX I

Wave Spectra

A theoretical analysis was performed in an attempt to explain mathematically the total process of generation, transmission, and reception of these pressure waves as typified by Figure 1. This theoretical approach is fundamental in nature as it provides understanding of the characteristics to be expected from signals so formulated. It is planned to present this original and extensive theory in a separate publication as yet undetermined.

Of interest, though, to those concerned with the physics is that a close examination of the experimentally obtained signals of Figure 1 show a high correlation with these theoretical findings. Pronounced characteristics that are readily apparent in Figure 1 coincide with the analytical results. This figure shows that between approximately 47 and 53 minutes from launch-time, the normal signal consists of relatively prominent groups of signal spectra having amplitudes clustered at positions located 48, 49, and 50.7 minutes from launch time. It may be further noted that the first group of spectra at 48 minutes is relatively small in amplitude, while the second two groups are contrastingly large.

This clustering and relative amplitudes are in close accord with the theoretical findings, which show an expectation that three major time periods of energies will be experienced in the reception of such pressure signals. The first period is mathematically predicted to consist of relatively low-level high-frequency transients. Then follows two major groups of pressure spectra considerably larger in amplitude than the initial transient group.

This analysis therefore shows, as experienced experimentally, that one should expect a first-arrival low-level response to the initial missile initiated energy spike. Then the two larger groups of pressure waves follow, as Figure 1 clearly illustrates.

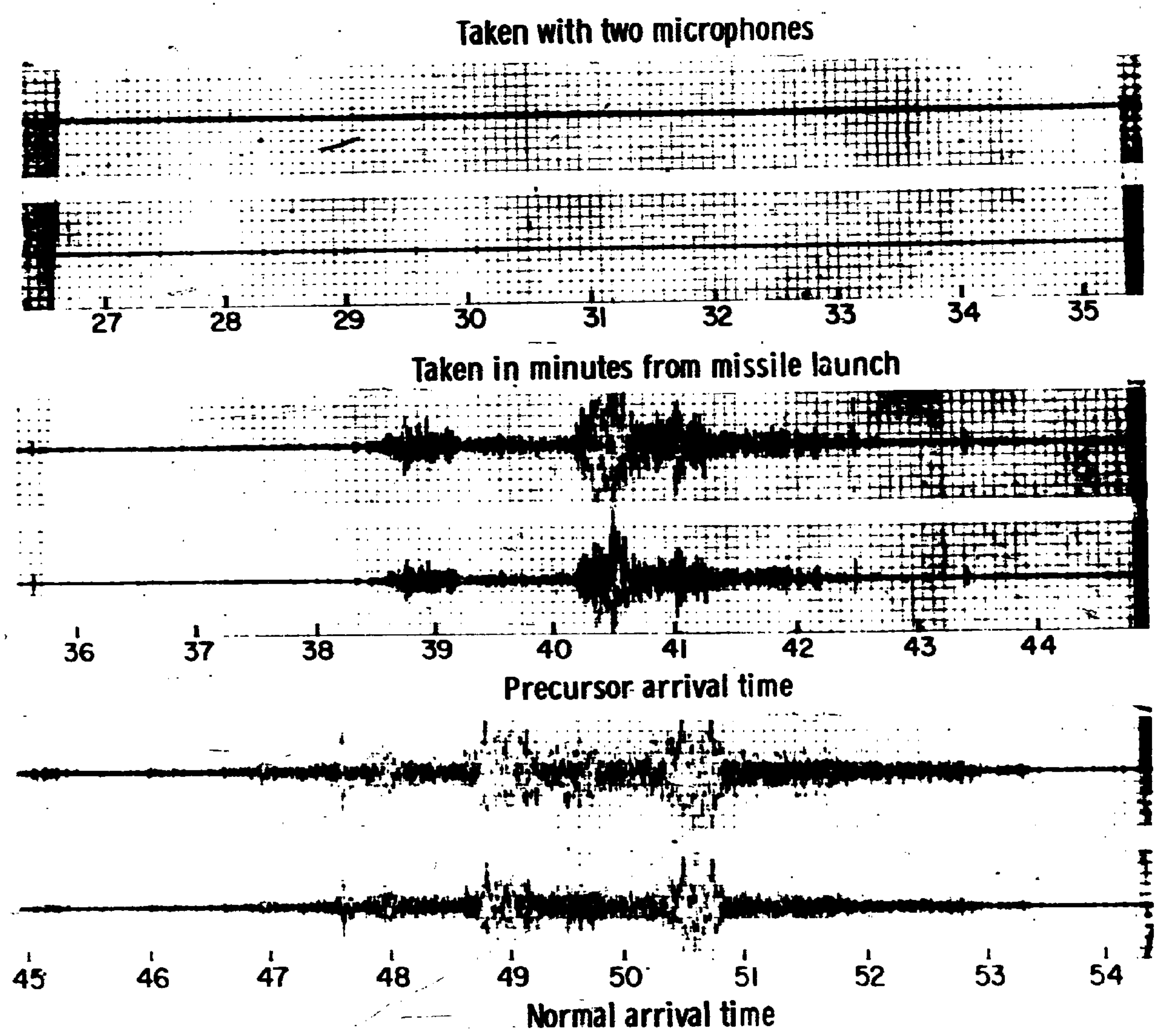


Fig. I. Infrasound 0.7 to 7 Hz Recording at Cape Hatteras of a Saturn-Apollo Launch From Cape Kennedy

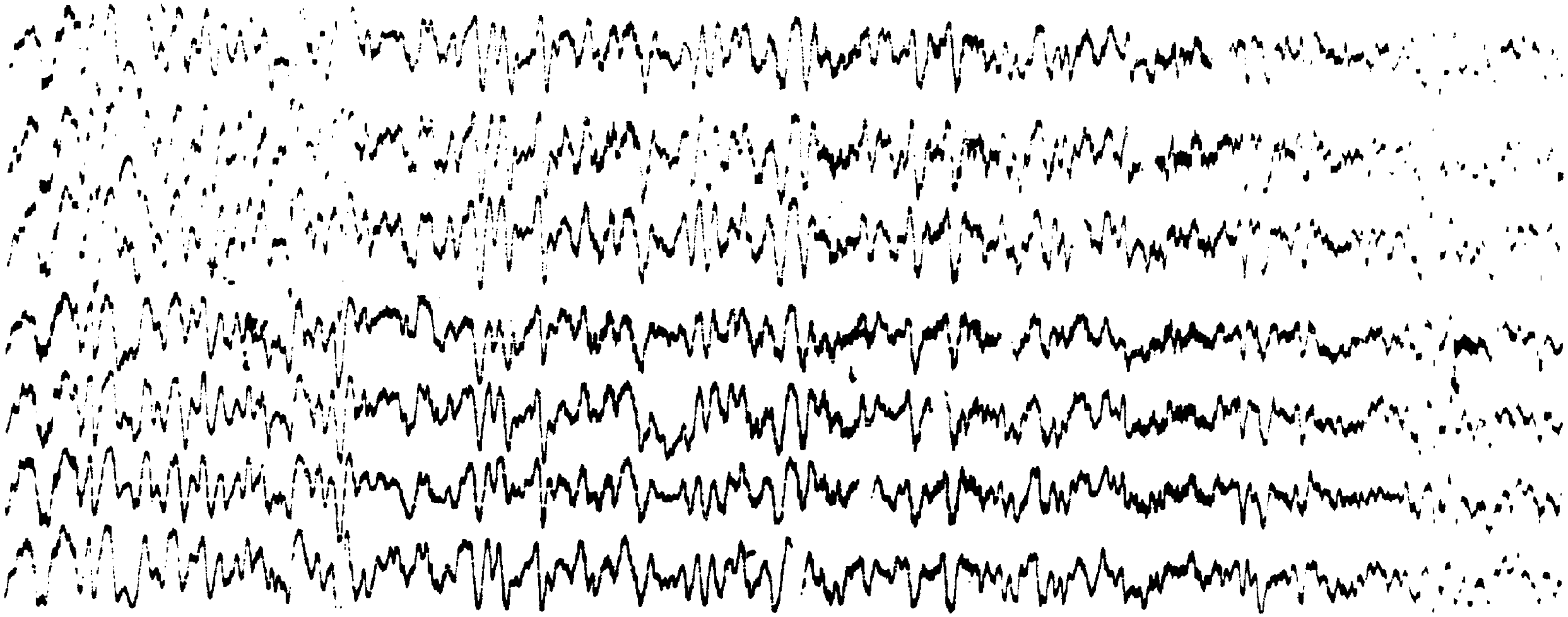


Fig. 2. An example of the correlated output from seven microphones in a broadside array.

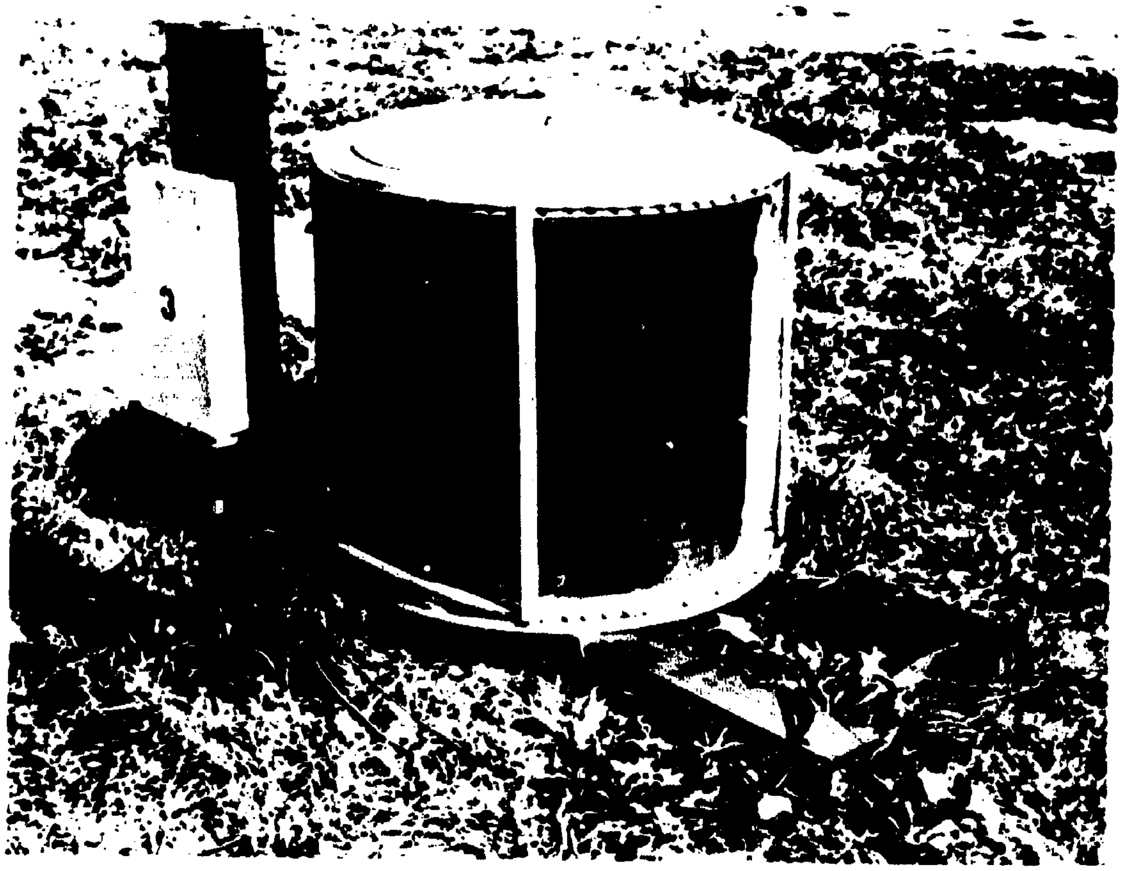


Fig. 3. An experimental design of an infrasonic wind screen inclosure.

Fig. 4. Experimental infrasonic broadside array of 20 microphones.

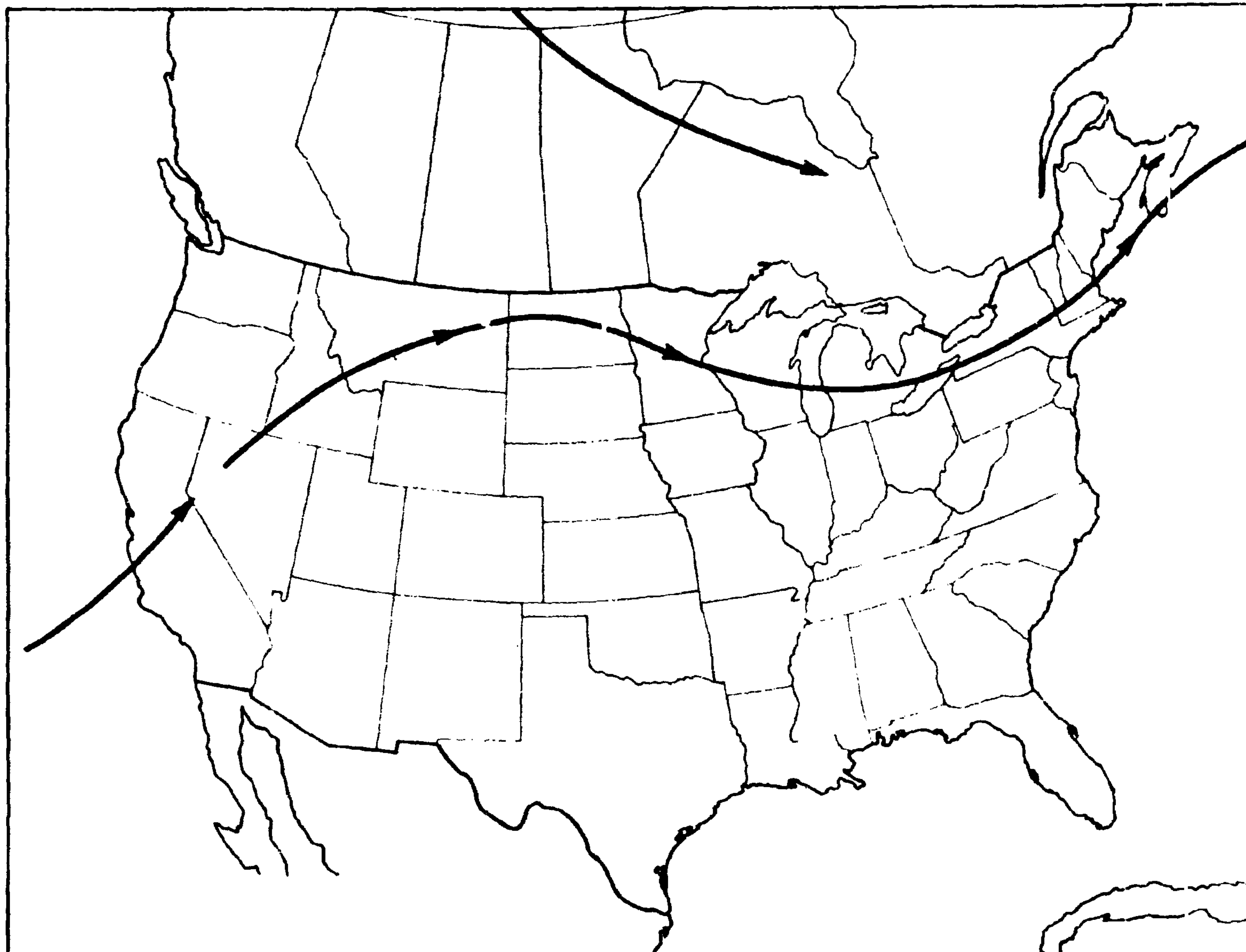


FIG. 5 TYPICAL SUMMERTIME JET STREAM PATTERNS

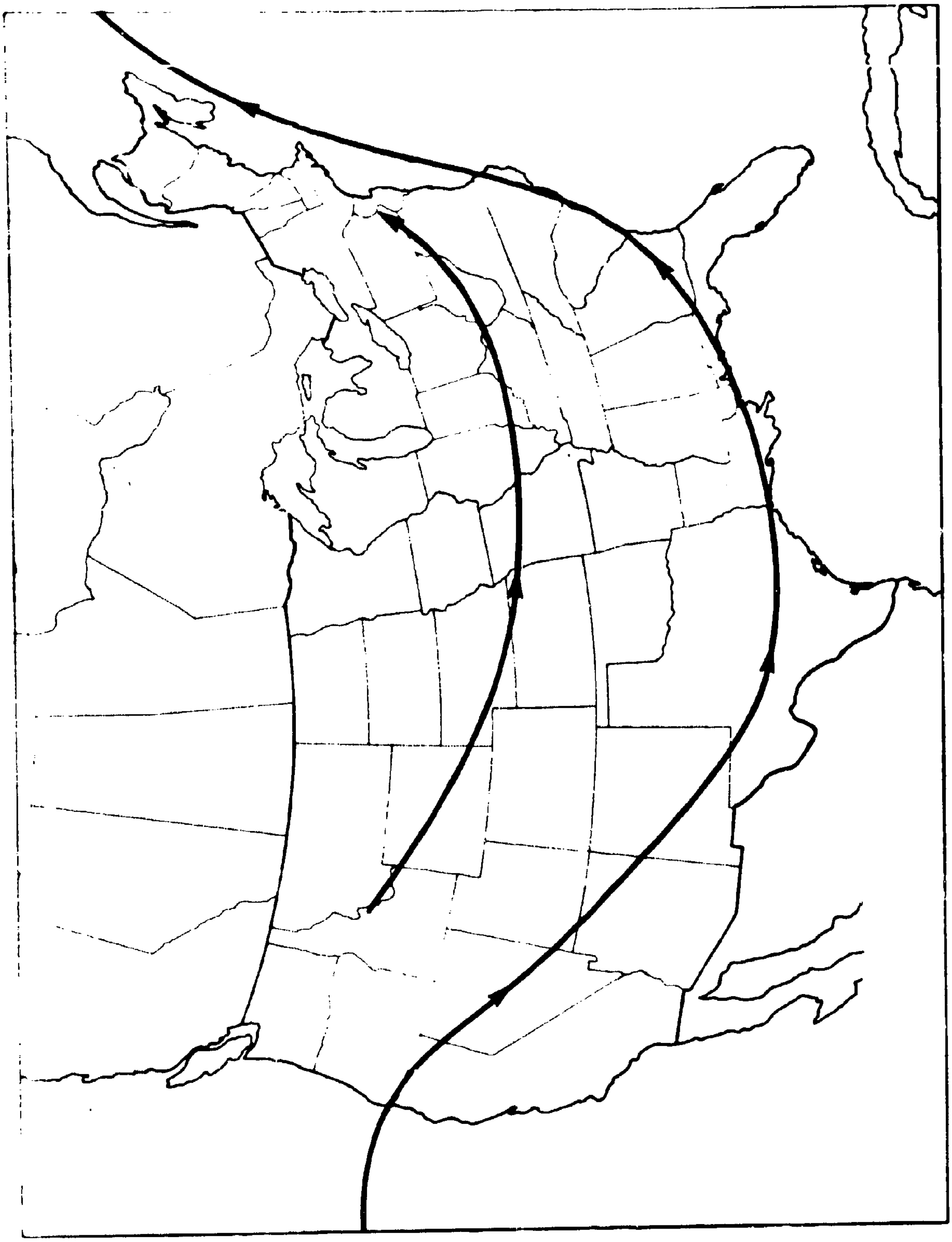


FIG. 6 TYPICAL WINTERTIME JET STREAM PATTERNS

DETERMINING THE FLIGHT RELIABILITY OF AN ANTITANK MISSILE WITH SIDE JETS

BY
R. G. CONARD AND N. R. RICH

U. S. Army Missile Command
Redstone Arsenal, Alabama

ABSTRACT

This paper deals with the probability of catastrophic failure and target miss of an antitank missile in flight resulting from failure of control jets to fire. When two failures occur in sequence, the missile deviates widely from the flight path but is recoverable. When three jets fail to fire in certain sequences, catastrophic missile failure will result.

A complete solution to this problem would involve the determination of the probability of failures of all possible combinations of successes and failures which would result in ground impact. It would consider the random distribution of the location of the target relative to launch position and the distribution of the missile about an average flight path.

A partial solution to this problem has been found in a system of "states," or intervals in the vertical plane. The recursive equations for the probability of lying in each state have been developed and tables of state probabilities for several values of the probability of success for one jet pair.

INTRODUCTION

The specific problem addressed in this paper deals with the probability of failure for a missile which employs jet pulses as a control mechanism and as a means for overcoming gravity.

In this system pairs of jet pulses fire sequentially as the missile, which maintains a fairly constant roll rate, flies toward the target. The pairs of jet pulses are so located about the center of gravity that little or no net torque is applied to the missile body (Figure 1). The firing position of the jets is in a near vertical plane. By varying the angle away from the vertical plane, a side component of thrust may be attained. This side force makes lateral corrections in the flight path upon command (Figure 2).

This article has been reproduced photographically from the authors' manuscript.

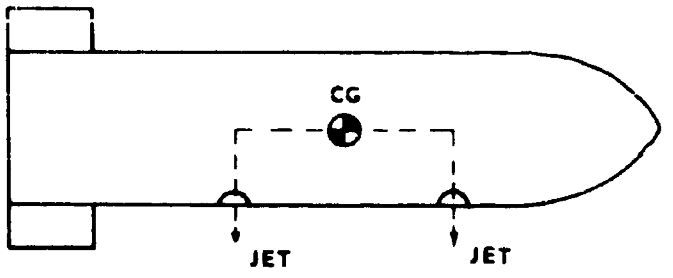


Figure 1. Location of Jets (Side View)

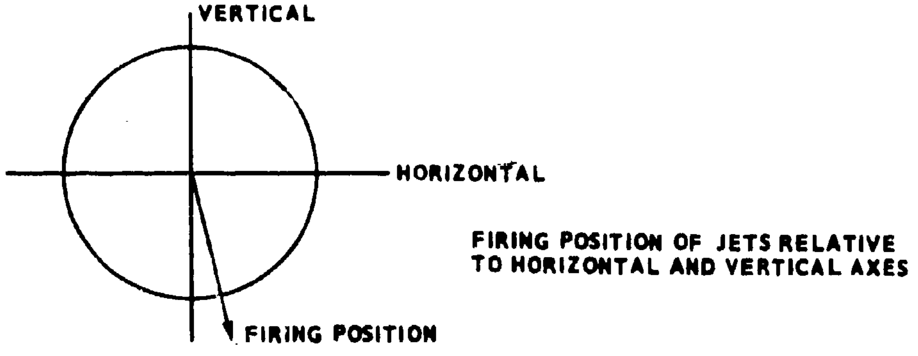


Figure 2. Location of Jet Pulse (Rear View)

The major concern is with the effect on the flight trajectory of a pair of jet pulses failing to fire at the correct time. Since the missile is acted on by gravity, it will drop immediately unless continually sustained by these jet pulses. The lateral dispersion will be affected also; this may or may not be serious depending on the location of the target with respect to location of missile when failure occurs and on the magnitude of such disturbing forces as cross wind. In this study, dispersion in the vertical plane alone is dealt with.

The missile is of a type which is continually commanded from launch to target. In this paper, it is assumed that failure problems do not exist in the guidance system, but occur only in the control mechanism.

So that the missile can be brought back to original flight path, the rate of firing increases as errors become larger. The time interval between jet firings is a discrete fraction of nominal firing intervals. The nominal firing rate will just balance gravitational forces, but a more rapid firing rate will force the missile to move upward.

OBSERVED FAILURE RATE

During the flight test program a known rate of failure has been observed. The failures have been in the circuitry so that either both jets of a pair fire simultaneously or both fail.

In the flight test program it has been observed that out of a total of M commands, L pairs have failed to fire. The estimated failure rate of the circuitry, therefore, has been L/M. There is no reason to believe that the rate will change unless the quality is improved at some additional cost or that the circuit is redesigned.

The effect on the trajectory has been investigated when combinations of failures have occurred in the flight program. Even though there has been a fairly high rate of component failure in the flight test program, the missile has not hit the ground. Although the missile exceeded a desirable control band for an interval of time, it would have missed the target only if the failure had occurred immediately before impact.

For this investigation two types of failures have been defined:

1) The missile deviates from the line-of-sight to the target by more than one unit. In this case it exceeds a desirable control band which will result in a miss if the deviation occurs just prior to impact.

2) The missile deviates downward three units and will impact the ground. This case is a catastrophic failure and can never recover or hit the target.

COMBINATIONS OF FAILURE

The first approach considered for this problem was that of determining the probability of certain combinations of failure.

If it is assumed that there are N jet pulse pairs, each with probability p of success and corresponding probability of failure $q = 1 - p$ and that the pairs fail independently of one another, the probability of no failures will be p^N. The following recursive equations hold:

1) Probability of no repeated failure (FF) in a sequence of N - (Table 1)

$$A_N = pA_{N-1} + pqA_{N-2}, \quad A_0 = 1, \quad A_1 = 1 .$$

In a system of N consecutive independent subsystems, each with probability p of success, A_N is the probability of no repeated failures.

P	.5000	.6000	.7000	.8000	.9000	.9500	.9750	.9900	.9950
N									
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	.7500	.8400	.9100	.9600	.9900	.9975	.9993	.9999	.9999
3	.6250	.7440	.8470	.9280	.9810	.9951	.9987	.9998	.9999
4	.5000	.6480	.7840	.8960	.9720	.9927	.9981	.9997	.9999
5	.4062	.5673	.7266	.8652	.9630	.9903	.9975	.9996	.9999
6	.3281	.4959	.6733	.8355	.9542	.9880	.9969	.9995	.9998
7	.2656	.4337	.6239	.8069	.9455	.9856	.9963	.9994	.9998
8	.2148	.3792	.5781	.7792	.9368	.9833	.9957	.9993	.9998
9	.1738	.3316	.5357	.7524	.9282	.9809	.9951	.9992	.9998
10	.1406	.2900	.4964	.7266	.9197	.9786	.9945	.9991	.9997
11	.1137	.2536	.4599	.7017	.9113	.9762	.9939	.9990	.9997
12	.0920	.2217	.4262	.6776	.9029	.9739	.9932	.9989	.9997
13	.0744	.1939	.3949	.6543	.8946	.9716	.9926	.9988	.9997
14	.0602	.1695	.3659	.6319	.8864	.9693	.9920	.9987	.9996
15	.0487	.1482	.3391	.6102	.8783	.9669	.9914	.9986	.9996
16	.0394	.1296	.3142	.5893	.8703	.9646	.9908	.9985	.9996
17	.0318	.1133	.2911	.5690	.8623	.9623	.9902	.9984	.9996
18	.0258	.0991	.2698	.5495	.8544	.9600	.9896	.9983	.9995
19	.0208	.0867	.2500	.5307	.8465	.9578	.9890	.9982	.9995
20	.0168	.0758	.2316	.5124	.8388	.9555	.9884	.9981	.9995
21	.0136	.0663	.2146	.4949	.8311	.9532	.9878	.9980	.9995
22	.0110	.0579	.1989	.4779	.8235	.9509	.9872	.9979	.9994
23	.0089	.0507	.1843	.4615	.8159	.9486	.9866	.9978	.9994
24	.0072	.0443	.1708	.4456	.8084	.9464	.9860	.9977	.9994
25	.0058	.0387	.1582	.4303	.8010	.9441	.9854	.9976	.9994

Table 1.

2) Probability of at least three failures in a row (FFF) in a sequence of N = (Table 2)

$$C_N = pC_{N-1} + qpC_{N-3} + q^3, \quad C_0 = C_1 = C_2 = 0$$

3) Probability of at least one pair of failure (FF) but no subsequence with three or more failures in a row = (Table 3)

$$W_N = 1 - A_N - C_N .$$

Trajectory simulation implies that when any combination of three out of four consecutive pairs or three out of five consecutive pairs of jets fail to fire in the right sequence, catastrophic failure results. Some combinations of three failures in six consecutive pairs result in marginal flight. A sequence of seven where the first, seventh, and any one other pair in between fail will not result in catastrophic failure. In fact, if every fifth jet pair failed to fire, the missile would recover but the dispersion about the desired flight path would be large.

This approach was abandoned since the important missile position could be found only by taking a particular combination of failures and running a trajectory simulation. The approach did not provide direct information on the probability of missile failure. For example, the missile would fail catastrophically if failures and successes alternated (SFSFSF...), but this case would be included in the calculation of A_N.

DISCRETE STATE STOCHASTIC MODEL OF VERTICAL DEVIATIONS

The next approach taken involved a simplified model of the vertical deviation from an average or nominal trajectory.

The average trajectory was computed by a least squares fit of the trajectory when all jets are firing in the proper sequence and at expected time intervals. The least squares fit established an average or expected trajectory about which the missile oscillated (Figure 3). In making the fit to the average trajectory the data immediately following the time when the failure occurs were discarded. The data were again used when the firing rate once more became normal.

At the time of firing of the N^{th} jet pair, the missile can be considered in one of seven "states." Each state is an interval in the vertical plane.

In a system of N independent consecutive subsystems, each with probability p of success, C_N is the probability of at least three failures in a row.

P	.5000	.6000	.7000	.8000	.9000	.9500	.9750	.9900	.9950
N									
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	.1250	.0640	.0270	.0080	.0010	.0001	0.0000	0.0000	0.0000
4	.1875	.1024	.0459	.0144	.0019	.0002	0.0000	0.0000	0.0000
5	.2500	.1408	.0648	.0208	.0028	.0003	0.0000	0.0000	0.0000
6	.3125	.1792	.0837	.0272	.0037	.0004	0.0000	0.0000	0.0000
7	.3671	.2151	.1023	.0335	.0045	.0005	0.0000	0.0000	0.0000
8	.4179	.2496	.1201	.0398	.0054	.0007	0.0000	0.0000	0.0000
9	.4648	.2826	.1377	.0461	.0063	.0008	.0001	0.0000	0.0000
10	.5078	.3141	.1551	.0523	.0072	.0009	.0001	0.0000	0.0000
11	.5473	.3442	.1720	.0585	.0081	.0010	.0001	0.0000	0.0000
12	.5837	.3730	.1887	.0646	.0090	.0011	.0001	0.0000	0.0000
13	.6171	.4006	.2050	.0707	.0099	.0013	.0001	0.0000	0.0000
14	.6479	.4269	.2209	.0768	.0108	.0014	.0001	0.0000	0.0000
15	.6762	.4521	.2366	.0828	.0117	.0015	.0001	0.0000	0.0000
16	.7022	.4762	.2519	.0888	.0126	.0016	.0002	0.0000	0.0000
17	.7261	.4992	.2669	.0948	.0135	.0017	.0002	0.0000	0.0000
18	.7481	.5212	.2817	.1007	.0144	.0019	.0002	0.0000	0.0000
19	.7684	.5422	.2961	.1065	.0153	.0020	.0002	0.0000	0.0000
20	.7870	.5623	.3102	.1124	.0162	.0021	.0002	0.0000	0.0000
21	.8041	.5816	.3241	.1182	.0171	.0022	.0002	0.0000	0.0000
22	.8198	.6000	.3377	.1239	.0179	.0023	.0003	0.0000	0.0000
23	.8343	.6175	.3510	.1296	.0188	.0024	.0003	0.0000	0.0000
24	.8476	.6343	.3640	.1353	.0197	.0026	.0003	0.0000	0.0000
25	.8599	.6504	.3768	.1410	.0206	.0027	.0003	0.0000	0.0000

Table 2.

In a system of N independent subsystems that are employed consecutively, each with probability p of success, W_N is the probability of at least one pair of consecutive failures but no more than two failures in a row.

P	.5000	.6000	.7000	.8000	.9000	.9500	.9750	.9900	.9950
N									
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	.3750	.2560	.1530	.0720	.0190	.0048	.0012	.0001	0.0000
3	.2500	.1920	.1260	.0640	.0180	.0047	.0012	.0001	0.0000
4	.3125	.2496	.1701	.0896	.0261	.0070	.0018	.0002	0.0000
5	.3437	.2918	.2085	.1139	.0341	.0092	.0024	.0003	0.0000
6	.3593	.3248	.2429	.1372	.0420	.0115	.0030	.0004	.0001
7	.3671	.3511	.2739	.1595	.0498	.0137	.0035	.0005	.0001
8	.3671	.3711	.3017	.1809	.0576	.0159	.0041	.0006	.0001
9	.3613	.3857	.3264	.2013	.0653	.0182	.0047	.0007	.0001
10	.3515	.3958	.3484	.2209	.0729	.0204	.0053	.0008	.0002
11	.3388	.4021	.3679	.2397	.0804	.0226	.0059	.0009	.0002
12	.3242	.4051	.3850	.2576	.0879	.0248	.0065	.0010	.0002
13	.3083	.4054	.4000	.2748	.0953	.0270	.0071	.0011	.0002
14	.2918	.4034	.4130	.2912	.1026	.0292	.0077	.0012	.0003
15	.2750	.3995	.4242	.3068	.1098	.0314	.0083	.0013	.0003
16	.2583	.3941	.4337	.3218	.1170	.0336	.0089	.0014	.0003
17	.2419	.3873	.4418	.3360	.1241	.0358	.0095	.0015	.0003
18	.2260	.3796	.4484	.3497	.1311	.0380	.0100	.0016	.0004
19	.2107	.3710	.4538	.3627	.1380	.0401	.0106	.0017	.0004
20	.1960	.3617	.4580	.3750	.1449	.0423	.0112	.0018	.0004
21	.1821	.3520	.4611	.3868	.1517	.0445	.0118	.0019	.0004
22	.1690	.3420	.4633	.3981	.1584	.0466	.0124	.0020	.0005
23	.1567	.3317	.4646	.4087	.1651	.0488	.0130	.0021	.0005
24	.1450	.3212	.4651	.4189	.1717	.0509	.0136	.0022	.0005
25	.1342	.3107	.4649	.4285	.1782	.0530	.0142	.0023	.0005

Table 3.

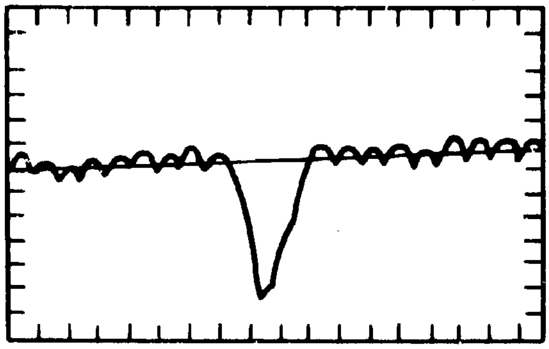


Figure 3. Vertical Deviation from Average Flight Path

If there are no jet failures, each pair will fire when the missile is in state 0 (Figure 4). The missile is considered to be in state 0 when the first pair is fired. Let X0(N) denote the probability of being in state 0 at the time of firing for N^{th} pair, then X0(1) = 1.

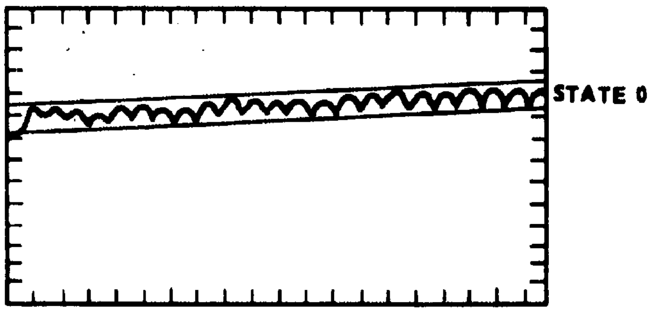


Figure 4. Nominal Trajectory (No Jet Failures)

When the first failure occurs, the missile will drop into state 2 (Figure 5). If the failure is followed by a success, the thrust of the jet pair does little more than arrest the missile in its fall; thus, the missile remains in state 2. The next success will bring the missile to state 1 and the next success brings the missile to state 0. For example, for the sequence S, S, F, S, S, S, the states will be 0, 0, 2, 3, 1, 0.

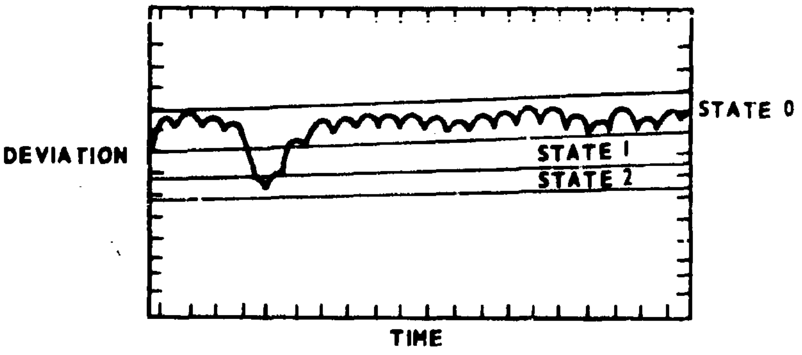


Figure 5. Vertical Deviation (One Jet Failure)

Two jet failures in a row after a series of successes will drop the missile into state 4 (Figure 6). If the next firings are successes, the missile will stay in state 4 one time and then climb up one state at a time. Thus, the sequence S, S, F, F, S, S, S, S, S, will result in the states 0, 0, 2, 4, 4, 3, 2, 1, 0. For the sequence S, S, F, S, F, S, S, S, S, S, S, as shown in Figure 7, the state sequence is 0, 0, 2, 2, 4, 4, 3, 2, 1, 0.

Three failures in a row will place the missile in state 6 (Figure 8). This is a captive state representing the catastrophic failure of hitting the ground. Once in state 6, the missile remains in this state.

The recursive equations for this model are given in Table 4. Here XI(N) is the probability of being in state 1 at the time of the N^{th} pair firing. These were developed through the following type of reasoning: The missile can be in

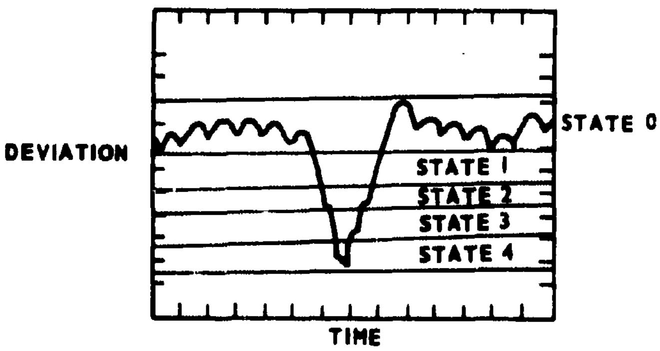


Figure 6. Vertical Deviation (Two Jet Failures)

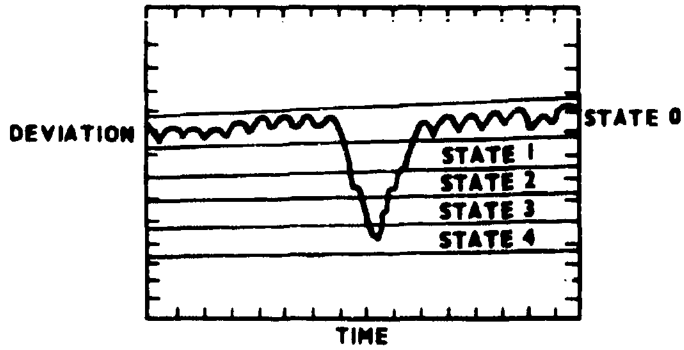


Figure 7. Vertical Deviation (Failure-Success-Failure Sequence)

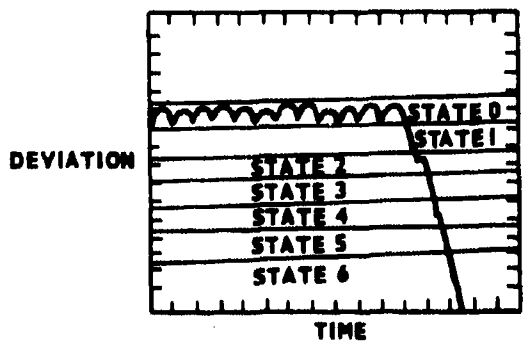


Figure 8. Vertical Deviation (Three Jet Failures)

(1) $X0(N) = pX0(N-1) + pX1(N-1)$

(2) $X1(N) = qp^2X0(N-3) + qp^3X1(N-4) + qp^4X2(N-5) + qp^5X3(N-6)$

(3) $X2(N) = qX0(N-1) + qpX0(N-2) + qp^2X1(N-3) + qp^3X2(N-4) + qp^4X3(N-5)$

(4) $X3(N = qX1(N-1) + qpX1(N-2) + qp^2X2(N-3) + qp^3X3N-4)$

(5) $X4(N) = qX2(N-1) + qpX2(N-2) + qp^2X3(N-3)$

(6) $X5(N) = qX3(N-1) + qpX3(N-2)$

(7) $X6(N) = X6(N-1) + qX5(N-1) + qX4(N-1)$

Table 4.

state 0 at the time of the N^{th} firing only if it were in state 0 or state 1 at the previous firing and that firing was successful. Thus, Prob (state 0 at time N) = Prob (state 0 at time N - 1) · Prob (success) + Prob (state 1 at time N - 1) · Prob (success), or $X0(N) = pX0(N-1) + PX1(N-1)$. It is assumed that the probability of success is independent of the state. The initial conditions for this set of recursive equations are given in Table 5. The values of the state probabilities for N = 1, ..., 25 and p = .50, .90, .95, .99 are shown in Tables 6 through 9.

States \ N	1	2	3	4	5	6
0	1	p	p^2	p^3	$p^4 + p^3q$	$p^5 + 2p^4q$
1				p^2q	p^3q	p^4q
2		q	$2pq$	$2p^2q$	$2p^3q$	$2p^4q + 2p^3q^2$
3					$2p^2q^2$	$4p^3q^3$
4			q^2	$3pq^2$	$4p^2q^2$	$4p^3q^2$
5						$2p^2q^3$
6				q^3	$4pq^3 + q^4$	$q^5 + 5pq^4 + 8p^2q^3$

Table 5.

The probability of catastrophic failure (hitting the ground) is the probability of being in state 6. If a target were of such a size as to cover states 0, 1, and 2 exactly, the probability of hitting the target would be the probability of being in any of these three states.

There are several drawbacks to this approach. The target might not cover an exact number of states so that the distribution within states is of importance. The range of the target should be considered to be variable; thus, the time between "times of firing" needs to be taken into account. As the pairs are fired, the number of available jet pairs decreases. Thus, if a failure occurs near the end of flight, the time until another unused pair is in position to fire is longer than it would be at the beginning of flight and the missile might drop further. For this reason, the model does not provide a uniformly good approximation to reality.

N	P	X0	X1	X2	X3	X4	X5	X6
1	.500	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	.500	.500000	0.000000	.500000	0.000000	0.000000	0.000000	0.000000
3	.500	.250000	0.000000	.500000	0.000000	.250000	0.000000	0.000000
4	.500	.125000	.125000	.250000	0.000000	.375000	0.000000	.125000
5	.500	.125000	.062500	.125000	.125000	.250000	0.000000	.312500
6	.500	.093750	.031250	.125000	.125000	.125000	.062500	.437500
7	.500	.062500	.031250	.125000	.062500	.093750	.093750	.531250
8	.500	.046875	.039062	.078125	.039062	.109375	.062500	.625000
9	.500	.042968	.023437	.050781	.050781	.085937	.035156	.710937
10	.500	.033203	.013671	.048828	.044921	.052734	.035156	.771484
11	.500	.023437	.013671	.043945	.026367	.041992	.035156	.815429
12	.500	.018554	.013671	.029785	.019042	.040527	.024414	.854003
13	.500	.016113	.009033	.021240	.019531	.031494	.016113	.886474
14	.500	.012573	.005981	.019042	.016235	.021362	.014526	.910278
15	.500	.009277	.005493	.016174	.010620	.017211	.013000	.928222
16	.500	.007385	.004943	.011596	.008087	.015289	.009368	.94 328
17	.500	.005164	.003479	.008682	.007446	.011871	.006698	.955657
18	.500	.004821	.002494	.007415	.006011	.008567	.005744	.964942
19	.500	.003658	.002166	.006088	.004230	.006889	.004867	.972099
20	.500	.002912	.001638	.004526	.003297	.005828	.003618	.977977
21	.500	.002375	.001348	.003479	.002853	.004536	.002706	.982701
22	.500	.001862	.001010	.002882	.002270	.003399	.002251	.986322
23	.500	.001436	.000847	.002223	.001672	.002723	.001848	.989149
24	.500	.001142	.000696	.001767	.001317	.002238	.001494	.991434
25	.500	.000919	.000524	.001376	.001098	.001748	.001076	.993255

Table 6.

N	P	X0	X1	X2	X3	X4	X5	X6
1	.900	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	.900	.900000	0.000000	.100000	0.000000	0.000000	0.000000	0.000000
3	.900	.810000	0.000000	.180000	0.000000	.010000	0.000000	0.000000
4	.900	.729000	.081000	.162000	0.000000	.027000	0.000000	.001000
5	.900	.729000	.072900	.145800	.016200	.032400	0.000000	.003700
6	.900	.721710	.065610	.145800	.029160	.029160	.001620	.006940
7	.900	.708588	.065610	.157464	.026244	.027702	.004374	.010018
8	.900	.696778	.076763	.153527	.024275	.030180	.005248	.013225
9	.900	.696187	.074401	.149393	.026572	.031686	.004789	.016768
10	.900	.693530	.071744	.149334	.029229	.030882	.004842	.020436
11	.900	.688747	.071744	.151620	.028219	.030345	.005314	.024008
12	.900	.684442	.074040	.150233	.027502	.030754	.005452	.027574
13	.900	.682634	.073222	.148726	.027894	.031036	.005289	.031195
14	.900	.680271	.072253	.148304	.028397	.030679	.005264	.034827
15	.900	.677272	.072036	.148432	.028041	.030443	.005350	.038422
16	.900	.674378	.072364	.147686	.027758	.030450	.005359	.042001
17	.900	.672069	.071963	.146891	.027765	.030427	.005299	.045582
18	.900	.669629	.071508	.146377	.027802	.030252	.005274	.049155
19	.900	.667023	.071252	.145994	.027634	.030196	.005279	.052708
20	.900	.664447	.071129	.145404	.027492	.030022	.005265	.056246
21	.900	.662020	.070831	.144798	.027406	.029931	.005235	.059775
22	.900	.659566	.070518	.144266	.027337	.029804	.005214	.063292
23	.900	.657076	.070258	.143766	.027218	.029684	.005200	.066794
24	.900	.654601	.070029	.143219	.027104	.029580	.005182	.070282
25	.900	.652167	.069760	.142668	.027009	.029475	.005160	.073758

Table 7.

[illegible]	[illegible]	X0	X1	X2	X3	X4	X5	X6
1	.950	1.[illegible]	.[illegible]	.[illegible]	.[illegible]	0.[illegible]	0.[illegible]	0.000000
2	.950	.950[illegible]	.0[illegible]	.250[illegible]	.0[illegible]	0.[illegible]	.000000	0.000000
3	.950	.902500	.000[illegible]	.095[illegible]	0.[illegible]	.002500	0.000000	0.000000
4	.950	.857375	.0[illegible]125	.090250	0.[illegible]	.007125	0.000000	.000125
5	.950	.857375	.042[illegible]	.135737	.004512	.00[illegible]25	0.000000	.000481
6	.950	.855231	.040725	.085737	.008573	.008573	.000225	.000932
7	.950	.851154	.040727	.089595	.008145	.008359	.000643	.001372
8	.950	.847200	.044[illegible]	.058[illegible]84	.007[illegible]	.008755	.000814	.001822
9	.950	.847193	.04[illegible]10[illegible]	.089307	.008271	.009[illegible]81	.000778	.002301
10	.950	.846[illegible]	.04[illegible]	.088[illegible]02	.008729	.009[illegible]	.000783	.002794
11	.950	.845861	.043646	.088776	.008641	.008963	.000826	.003284
12	.950	.845[illegible]	.044117	.088640	.008576	.009[illegible]	.000846	.003773
13	.950	.844691	.04[illegible]	.088504	.008616	.009043	.000839	.004266
14	.950	.844293	.043[illegible]40	.088463	.008677	.009026	.000838	.004760
15	.950	.843822	.043417	.088439	.008659	.009014	.000843	.005253
16	.950	.843[illegible]62	.043751	.088434	.008644	.009015	.000845	.005746
17	.950	.842948	.043[illegible]1	.088373	.008645	.009016	.000843	.006239
18	.950	.842[illegible]	.043697	.088331	.008649	.009010	.000842	.006732
19	.950	.842111	.043475	.088297	.008643	.009004	.000843	.007225
20	.950	.841656	.04336[illegible]	.088252	.008637	.009000	.000843	.007717
21	.950	.841271	.043339	.088206	.008633	.008997	.000842	.008209
22	.950	.840855	.043[illegible]1[illegible]	.088162	.008630	.008992	.000841	.008701
23	.950	.840437	.0437[illegible]	.088119	.008626	.008987	.000841	.009193
24	.950	.840017	.04377[illegible]	.088075	.008621	.008983	.000841	.009684
25	.950	.839604	.043751	.088031	.008617	.008978	.000840	.010176

Table 8.

N	P	X0	X1	X2	X3	X4	X5	X6
1	.990	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	.990	.990000	0.000000	.010000	0.000000	0.000000	0.000000	0.000000
3	.990	.980100	0.000000	.019800	0.000000	.000100	0.000000	0.000000
4	.990	.970299	.009801	.019602	0.000000	.000297	0.000000	.000001
5	.990	.970299	.009702	.019405	.000196	.000392	0.000000	.000003
6	.990	.970201	.009605	.019405	.000388	.000388	.000001	.000007
7	.990	.970009	.009605	.019596	.000384	.000386	.000005	.000011
8	.990	.969819	.009795	.019590	.000381	.000390	.000007	.000015
9	.990	.969812	.009791	.019583	.000385	.000393	.000007	.000019
10	.990	.969812	.009786	.019583	.000390	.000393	.000007	.000023
11	.990	.969804	.009786	.019589	.000390	.000393	.000007	.000027
12	.990	.969795	.009792	.019588	.000390	.000393	.000007	.000031
13	.990	.969791	.009791	.019588	.000390	.000393	.000007	.000035
14	.990	.969787	.009791	.019588	.000390	.000393	.000007	.000039
15	.990	.969783	.009791	.019588	.000390	.000393	.000007	.000043
16	.990	.969779	.009791	.019588	.000390	.000393	.000007	.000047
17	.990	.969775	.009791	.019588	.000390	.000393	.000007	.000051
18	.990	.969771	.009791	.019588	.000390	.000393	.000007	.000055
19	.990	.969768	.009791	.019588	.000390	.000393	.000007	.000059
20	.990	.969764	.009791	.019586	.000390	.000393	.000007	.000063
21	.990	.969760	.009791	.019583	.000390	.000393	.000007	.000067
22	.990	.969756	.009791	.019588	.000390	.000393	.000007	.000071
23	.990	.969752	.009791	.019587	.000390	.000393	.000007	.000075
24	.990	.969748	.009791	.019587	.000390	.000393	.000007	.000079
25	.990	.969744	.009791	.019587	.000390	.000393	.000007	.000083

Table 9.

One of the limitations of this approach is the difficulty of relating the position of the missile in one of the various states to its position along the trajectory since the rate of firing of the side thruster is not always a constant. In order to compensate exactly for gravity there must be a predictable amount of impulse imparted to the missile during a given time interval. If one side jet pair fails to fire at the proper time and the missile develops a vertical acceleration due to gravity, the acceleration must be corrected and the missile brought back to its appropriate flight path by more rapid firing of the side thrusters.

If the success-failure sequence is known, the number of successful side jet firings required to bring the missile back to its original flight path can be determined by flight simulation. For the simpler cases this has been done and a time relation between state position and the required number of successful firing results has been determined explicitly. The more complicated sequences would present some difficulty especially if the supply of available side jets is exhausted. On the average, however, a missile in a given state whose vertical acceleration has been arrested will have experienced a predictable number of successes within a given time.

QUESTIONS TO BE CONSIDERED

The following questions are submitted for consideration by the panel.

1) The distribution of the missile deviation about an average flight path will involve the dispersion of a properly functioning missile as well as the dispersion resulting from jet failure. A technique is required for combining the two sources of error.

2) Since the target is at a random position relative to the launcher, the probability of hit may be related to the range of the target. How should the target range be included as an error source?

3) Toward the end of flight, the number of jets to be fired becomes limited. Therefore, if a failure occurs late in the flight, there might be empty rows of jets. This would result in a delay in firing after the failure until a pair of jets would be in the proper position to be fired. How should this problem be handled?

A PROBABILITY APPROACH TO CATASTROPHIC THREAT

Clifford J. Maloney*
Bethesda, Maryland

I. INTRODUCTION. Prior to World War II statistical methods were viewed as a body of techniques appropriate to scientific research and to a limited range of other activities, primarily insurance. The war and the immediate post war years saw an enormous expansion of the application of statistics to quality control of incoming supplies, operations research as a replacement for trial and error in the choice of operational organization and technique, and sample surveys as a substitute for complete canvass for obtaining socio-economic data needed in decision-making. Developments more purely mathematical in nature, including linear programming, development of the electronic computer, finite mathematics and the like served to help foster an increasing mathematical, probabilistic, and statistical approach to decision-making in all aspects of human activity.

Perhaps the one aspect which has so far not attracted the attention of scholars has been the anticipation, prevention, and/or amelioration of unexpected but costly contingencies; an activity the businessman has long known as "putting out fires." But despite the fact that, superficially at least, "every case is different," the very fact that in the aggregate these occurrences are both frequent and costly, suggests that at the very least we can attempt to codify the techniques by which the successful practitioners cope with these pheonomena, however diversified, and that a possibility as well exists that some underlying structure will be discovered.

II. THE NATURE OF CATASTROPHE. The Statistical Department of the Metropolitan Life Insurance Company defines a catastrophe as any accident in which 25 or more lives are lost. Two incidents will show that an event may be a disaster despite the fact that no loss of life occurred or was even threatened. The first was the Pueblo incident where the chief cost was the severe, and continuing, but unmeasurable contribution to the threat to peace. The second was the public reaction to the characterization of a military briefing as a "brainwashing." One suspects that, two years later, the reaction would be very different. A first attempt to list the characteristics of a disaster is given in Figure 1. The ensuing six figures give a few ill-chosen examples under each of six disaster characteristics to which the reader can undoubtedly add an indefinite number of more suitable examples.

My claim that while, in fact, unexpected the disastrous event was "predictable" can be illustrated by the seizure of the Pueblo. All parties recognized that such an act was conceivable. The difference lay exclusively in their assessments of its likelihood. All I mean by

*Views expressed herein are those of the author and not to be construed as official.

predictable is that the catastrophic event can be visualized prior to occurrence. Contamination of the moon by earth microbes, or of the earth by organisms returning with the astronauts is another such case. Earth rupture, tsunami, or earthquakes are all conceivable consequences of underground nuclear tests. It is unnecessary to argue that the inconceivable never happens in order to make it clear that most catastrophes are of an ordinary garden variety; that their catastrophic nature lies in their (1) severity and (2) unexpectedness, not in their absolute novelty.

That catastrophe severity is as much psychological as real, may not be at once apparent. Such loss (other than one ship and the suffering of the men and their families) as was sustained in the Pueblo incident, lies in the threat to world peace - i.e., is psychological. The Dugway sheep kill created a great public outcry - yet the anticipation that several workmen will lose their lives when the Washington, D. C. subway is built, won't raise an eyebrow. The brainwashing episode was an extreme example.

III. DECISION-MAKING DOMAINS. To decide to ignore or to anticipate a threat is to make a decision, and decision-making is one of the more recent and more prestigious applications of probability. If, as claimed, catastrophic threat has not hitherto been recognized as an appropriate application of decision theory, there must be something about the material that conceals its appropriateness for the technique. We can at once expose this unconscious assumption, show how to remove it, and show the absolute necessity of doing so by reviewing the circumstances in which the application of probability is currently fully recognized. First, partly as a horrible example, partly because all problems not treated with a formal algorithm are necessarily so treated, and partly to show the absolutely fundamental role of probability in all of life's decisions, the universally applied method of common sense will be listed as a specific method and discussed. Its characteristics are given in Figure 8. Many, if not most, of life's decisions hinge on an estimate of probability. To portray this rule in a military context, I have extracted certain sentences from the chapter on the Anzio Campaign from the book COMMAND DECISIONS, published by the Office of the Chief of Military History in 1959, and present them in the appendix. Figure 9 is a map of the theater involved. By reading these extracts, which are limited to only those sentences in which a probabilistic concept is essential, in order one can gain a picture of the campaign almost as complete as by reading the full text.

Two further points should also be clear. The various generals assessed the probabilities very differently. In particular, General Mark Clark feared over-precipitate advance because of his one-time experience at Salerno. The English authorities seemed to feel that the German Cassino Line would crumble the moment the Allies had gained a foothold. No Allied commander seems to have contemplated the possibility of an unopposed landing - which in the event is what happened. What I seek to infer from this example is that up to the time of World War II, no formalized procedure existed for dealing with the probability component of command decision-making and that that lack proved costly.

The application of probability considerations first occurred, and by now is well established, in the situations which grew out of, or which can be logically reduced to, games of chance. In practice, these cases arise in designed experiments or in sample surveys where the sample is drawn by "probability sampling." This situation is characterized in Figure 10.

In a closely related but distinguishable situation, characterized in Figure 11, the sample is not drawn from a preformed population but from a growing one. Both of these cases (Figures 10 and 11) involve probabilities and sample sizes in a middle range.

In the next situation (Figure 12) we are dealing with a case where, again, probabilities are, or at lease often are, in the middle range, but sample sizes are usually not even defined. During World War II, the Weather Bureau developed a so-called 30-year series of Northern Hemisphere maps to enable military forecasters to make forecasts by matching the current weather map with an earlier situation, on the assumption that the succeeding weather should also be a repetition. Unfortunately, matching is never exact. At this point a new factor is brought into play, which goes back to the earliest recognition of a role of chance assessment in decision-making, but which was relatively neglected in the century preceeding World War II, while an objective foundation for probability seemed to be proving adequate. This new factor was the concept of personal probability, not as a quantity having external reality, but as a measure of subjective mind state whether shared by several minds or peculiar to one. It will be argued below that this factor has been incorrectly apprehended. Here it is sufficient to note that on this foundation a super-structure for dealing with forecasting, technological and otherwise, has been elaborated and is being increasingly widely adopted. A good introduction is afforded by Bright (1968) and the references there included.

Finally, we come to the situation contemplated in this paper. Like the case above, sample size is essentially non-existent. But here the probabilities are so low that no hope exists of accumulating a fund of cases, sufficiently "similar" to provide more than the most tenuous basis for probability quantitation even as a belief state. This is the case described in Figure 13. The essential difference between this case and that of Figure 12 is that, whereas in the latter, probabilities are fairly high, here they are low. For example, stock market, weather or weapons system forecasting cannot successfully be based on simple sample statistics, as for instance is done when the number of telephones or children per household, or the weight gain due to a ration additive is assayed in a controlled trial; still the market opens, the next day comes and brings with it an after the fact verifiable weather outcome, and whatever weapon system is chosen for development, five to ten years justifies the decision or reveals its invalidity. But catastrophic threat may or may not come to pass. The air defense of the United States against a bomber attack by the Soviets was incredibly expensive. Was it a success? No such attack occurred, but would one have had one if the defense been less effective or less costly? We will never know. The current debate over ABM deployment would not continue a day if all concerned were agreed that within the decade of the seventies one or more technologically effective nuclear missiles were highly likely to be launched against one or more of our cities. The ABM proponents, quite as much as the opponents, not only hope that the event will not occur, but assess its likelihood as low. The disagreement is entirely over how low is low enough.

IV. PROBABILITY FOUNDATIONS. Three times in the past the development of a new foundation for quantitative probability has resulted in a major success in the increase in understanding and control of the real world. The two "objective" foundations, the a priori or necessary of Cardan Galileo, Fermat and Pascal, and the frequency foundations of Graunt and Halley, developed about the same time, proved adequate as a foundation for all applications of probability prior to World War I. They were abstracted to a formal mathematical calculus at the end of that period by Kolmogoroff and others.

For the most part the necessary and the frequency foundation for probability are complementary. There is, however, one aspect in which they give conflicting counsel, the problem of outliers. On the frequency definition an outlier is meaningless - for that approach takes experience as given. On the contrary, the a priori or necessary approach rejects experience whenever it is in conflict with doctrine. Adherents of both approaches, being sensible men, adhere neither to the one extreme nor to the other. Nor, indeed, do they in practice achieve a comfortable compromise.

As suggested above, notions of "degree of belief" as a foundation of probability theory, while studies in the years following World War I, were not widely employed until after World War II. Most adherents of personal probability adopt a concensus form of the approach. It is not what one believes, but what one ought to believe that matters. DeFinetti and Savage, however, go all the way and allow everyone his own private degree of belief, divorcing the concept entirely from external world events.

A fifth foundation for the concept of probability underlies, if somewhat subconsciously, most work on curve fitting and indeed comes close to formal enunciation especially in the older least squares literature. Here every value of every quantitative property is assumed to be precisely determined, but as a function of a literally infinite number of independent or predictor variables, say

$$X = \alpha + \beta + \sigma + \delta + \zeta + \ldots$$

One achieves a chance structure by neglecting most of these latter. Then the precise value of the predicted variable depends on the contribution of the neglected variables, which being unknown makes the predicted value a chance one. Each of the other theories can be obtained by making assumptions about this formula. The a priori definition assumes that the relative frequency of possible values of X can be inferred because the values of the independent variables are known, whenever individually important. The frequency definition conversely finds the values only of a fraction of the parameters, age, sex, occupation and so on in the case of mortality schedules ascertainable, but a number of essential parameters remain unknown, so that observations on X, as well as calculations on the known right hand terms, are required. Either subjective theory would differ from the a priori approach only in that the values of the parameters and the form of the function would be apprehended intuitively rather than be calculated from theory or inferred from sampling experiments. How does a batter hit a ball is the classic example to distinguish between the rational, algorithmic approach to nature and the intuitive subconscious unanalyzed approach. The batter seldom regards a graduate degree in mechanics as essential for success - though he does indeed bear in mind all the identifiable admonitions which can be derived either from theory or experience.

Personal probability is to be distinguished from objective probability not because it is an entirely different - an unrelated - species, but because it stresses a component always present when a mind contemplates the external world. The mind never knows the world, it has only beliefs about it. These beliefs are a substitute, or better, an alternative for objective measures. In some contexts, the objective criteria prove more successful, in others the subjective judgments. These beliefs can have little value if not verified in practice, and verified in just the way that objective deductions are - by the observed outcome. If the odds makers pick the wrong team to win the World's Series, all agree that the choice was an error. One is entitled to his opinion - but only before the game. The several bases for the concept of probability are set out in Figure 14.

V. A NEW VIEW OF PROBABILITY. It was remarked above that the requirements of supplying a foundation for decision theory in circumstances where evaluating parameters by sampling of many indistinguishable elements is not available, has since World War II been furnished by the concept of probability as "degree of belief." It was further asserted that the true nature of personal probability is not, as often thought, something apart from objective probability. There is only one probability. That is the success ratio in a series of trials. A logical complication arises from the fact that the series is necessarily infinite and therefore necessarily conceptual rather than actual. It may be estimated in each of three ways. The first way is the a priori or necessary approach.

If the possible results can be enumerated, if they possess an adequate degree of symmetry, and if they are independent, we apply the principle of insufficient reason. This approach has always been intuitively appealing, logically intractable, and limited in application. The device of just trying or observing a large number of trials, partially overcomes this latter difficulty, shares in a different way the intuitive appeal, but introduces its own logical difficulties. The degree of belief approach possesses the prime advantage of always being applicable, but only to the perceiving mind, never to the real world. For this latter to occur we must appeal to a method of verification. For example, consider a weather forecaster. Each newscast he presents a "fearless forecast" preceded by an explanation of why his last forecast failed. Weather forecasts are made on the basis of much data, much theory, historical records, or Farmer's Almanacs. The basis is irrelevant. The verification is the thing. If the forecasts are verified, the forecaster is a success. The sequence of trials is thus the successive forecasts periods. The successes are the verified forecasts, all others are failures. The process is just as objective as was Graunt's birth series. That the forecaster used his degree of belief in formulating his forecast is as incidental as the use by a batter of his in deciding to swing at a pitched ball. Each man functions as a measuring instrument. A person is often superior to a machine in such functions.

The catastrophic threat problem does introduce a complication. In most, if not all, realistic applications of personal probability the verification is not long in coming. That is why people forecast. But a catastrophic threat is necessarily a rare event, if it happens at all. Just as we cannot, before the event, accumulate an adequate fund of experience upon which to base an estimate of probability, so we cannot after the verification step accumulate a sufficient fund of verified or disproved forecasts to establish a success ratio in a trial sequence known to be homogeneous in the probability sense. Fortunately, the occurrence compensates for its severity by its rarity. The new element which the present analysis seeks to supply to the contribution of probability to decision-making is a new technique of verification.

For this purpose consider Figure 15. Here we have listed not one stochastic sequence, but three labelled as conditions A, B, C. A, B, and C could be threats of aggression by three different nations, or the threat of war, of famine and of a major epidemic. Each represents what would usually be chosen to characterize one series of trials as ordinarily described in discussions of probability. But the types of catastrophic threat are not limited; they are legion. Suppose we employ the methods of personal probability to assign "degrees of belief" to occurrence of each. Further, let this be done not by one individual, or by one forecasting team, but by many, each applying whatever "betting system" is most attractive. We would get as a result a two-fold matrix of assigned probabilities such as is diagrammed in Figure 16.

While it is true that the probabilities subjectively assigned to the contingency that a catastrophe of specified nature would occur in a specific interval of time is low, and in general varies as between the types of contingencies involved, the possible contingencies are legion, so that it is to be expected that on any given estimation procedure a fairly high number of possible catastrophes would be assigned the same or nearly the same value of the probability of occurrence in a specific time frame. In Figure 17, this is indicated by assigning probabilities not to individual catastrophic contingencies, but to classes characterized by the property that within a class, each individual catastrophe is regarded as equally likely. The grouping may well be different as between rows. Hence, the columns do not refer to the same contingencies in every row. The letter k in each cell represents the number of contingencies all assigned the same probability of occurrence by the technique of that given row.

As each period of observation passes, the occurrence of the various types of catastrophe would be noted. The subjective probabilities assigned by any one procedure (in any one row) would be verified or refuted according as the empirical success ratio was sufficiently close to the ex ante assigned probability or not. This empirical success ratio would be calculated in a slightly different manner from that utilized at present, where it is assumed that every sequence of trials is studied in isolation and without regard to what is happening in any other. Let the success ratio at a certain observation period be n/m; where n is the number of catastrophes, all having been assigned the same probability, which have so far occurred and m is the product of the number of such catastrophes by the number of observation periods. Here k equals the number of catastrophes grouped as having the same subjective probability of occurring, and p is the probability itself. Let h be the number of catastrophes of this class which occur in the next observation period. Of course, h will almost always be zero, rarely be unity and almost never be greater. Then the success ratio at the end of this period becomes

$$\frac{n + h}{m + k}$$

The essence of this procedure is that we are judging, not the success in assigning subjective probabilities to specific types of catastrophes, but the betting system itself (at this level of probability). If the probability assignment for one class of catastrophe is verified, our confidence in the assignment of all classes is strengthened.

So far we have pooled experience in assessing the efficacy of a particular "betting system," i.e., method of assigning subjective probabilities, but only at a specific level (or limited range of levels) of the assigned probability. As first shown by Karl Pearson, we can use chi-square or some alternative approach to get a combined test of the procedure irrespective of level.

Particularly in a medical environment the need to ensure a very high level of safety for products produced for human consumption has received increasing attention. The Polio and Thalidomide incidents are instances. The cigarette and cyclamate episodes provide an illustration of variation in response to evidence. These cases have been studied (so far as the author is aware) by extrapolation of conventional statistical techniques from regions of relatively high probability (so that evidence is attainable) to the region of interest where probability of occurrence is extremely low - one in a million exposures or less. The present paper seeks to provide a procedure for catastrophes where even this extrapolatory technique is unavailing, but there is no reason why, when it is, that the evidence from both approaches should not be pooled.

VI. THE COST FACTOR. Costs, like threats, are some real and some imaginary. A perfect system of probability assessments would still not be adequate for decision-making, if costs are ignored, except in those instances where the probability concerned is negligible.

There is a vague appreciation that in a military context "negligible" probabilities are a trap. Defenders, particularly budget officers who are to fund protection from such "impossible" threats, or competitors for those funds inevitably dismiss such probabilities as negligible, yet often in fact, just that stratagem will be selected by attackers in the very knowledge that it will be unanticipated by the defenders. Instances are the Pearl Harbor attack; the choice of land over sea approach at Singapore; the Black Forest end run of the Maginot Line; the "post season" sea assault at the Battle of Hastings. Two famous examples from the history of mathematics describe the failure of Saccheri, and the success of Hamilton from a confrontation with the unthinkable.

As this paper treats catastrophic threat the costs are by assumption high. But a closer estimate is needed. Perhaps the outstanding characteristic of disaster costs is that the ex post[2] estimates of costs "always" outweigh the ex ante[1] cost estimates, and by a huge factor. This is recognized by every parent who tries to get his teenage offspring to cross a college campus or to drive the family car with due caution. The military meets this problem in stark terms when it tries to indoctrinate habits of safety in new recruits. Live ammunition in field exercises represents one attempt to secure credible reality. However, no adequate technique has been found.

---

[1] before the observation period.
[2] after the observation period.

But while a closer approximation of ex ante to ex post costs assessments by new troops would reduce the incidence of personal tragedy, and shorten the battle seasoning of troops, the wide discrepancy inevitably existing in assessing the costs of catastrophes, is the essence of the catastrophic threat problem. Indeed, in practice all or nearly all the discussion will be found to turn on an estimation of the probabilities, and little of it on assessment of the costs, though a great deal of attention is devoted to deprecating the costs of prevention or of protection predicated on this presumed negligible risk. With a one-sided consideration of both costs and probabilities, there is little chance of efficient decision-making.

The instinctive impulse of the mathematically, or philosophically minded is to generalize the problem and then return to the specific example - in our case the unreality of ex ante estimates - on the basis of supposedly clearer insight. To estimate is to make a decision, and decision-making is intrinsically emotional. This point was graphically demonstrated by Koehler in experiments during World War I, Figure 18.

If a chicken is placed close to a short stretch of fence and a handful of grain is placed on the other side close to the fence, but beyond reach through it, the chicken will excitedly press against the fence and never step back to seek a way around it. If the grain is moved farther away, the emotional attraction will be reduced, and the hen will have a better chance of solving the problem.

So true is it that decision-making is an emotional process that advertising and political discussion are directed almost exclusively at emotional, and only incidentally at logical components of issues. In consequence persons who will make decisions are selected on the basis of emotional, not rational, characteristics. So widespread is this process that it almost becomes a "first law of management" that, when it comes to making decisions "those with the power lack the knowledge, those with the knowledge lack the power." In consequence, the "on tap" become the "untapped."

This lack of knowledge has not merely been recognized, but has been deplored through all of history. Even so, courts of inquiry, inquests, grand juries, staffs, study groups and a host of other techniques, have evolved to supply knowledge deficiencies of those in power. However, in a power rather than information dominated environment, all such devices tend to be ineffective because they tend to be ignored. Where power makes decisions, only power can influence decisions. Where decisions are out of touch with reality because information is ignored, and countervailing power is laking feedback from consequences is avoided. Figure 19 lists certain principles which seem applicable.

Ascendency of emotional over informational factors in decision-making leads to a widely recognized, widely deplored, oscillation in efforts at averting or dealing with catastrophic threat. The soldier knows this oscillation in the contrast between his hero's role in times of threat. and scapegoat's role in seeming peace. The tin-roof parable is known to all, but heeded by few. The technique outlined in this paper represents only what <u>could</u> be done, but not what <u>will</u> be done, until the forces producing this violent oscillation are rendered impotent. A single incident at this Design Conference will illustrate the situation. One talk at the Conference was given by Dr. Condon, in charge of the safety program for the moon landings. After explaining that the success of the safety program was due, in large part at least, to his refusal to adopt a "good enough" attitude, the speaker conceded that the program was thereby made costly and commented that, in the days ahead there will be "pressures on us in the way of costs."

This is the universal story of the fight against catastrophe. The expense ensures success; the success breeds lack of support. It will always be thus where decision-making is a budgetary process.

REFERENCES

Bright, James R. ed. [1968]. Technological Forecasting for Industry and Government. Prentice-Hall, Englewood Cliffs, New Jersey.

DeFinetti, Bruno [1937]. Foresight: Its Logical Laws, Its Subjective Sources. Annales de l'Institut Henri Poincare, Vol. 7. Reprinted in Studies in Subjective Probability, Kyburg & Smokler, eds. Wiley.

Quade, E. S. ed. [1964]. Analysis for Military Decisions. The Rand Corporation, Santa Monica, California Report No. R-387-PR.

Savage, Leonard J. [1954]. The Foundation of Statistics. John Wiley & Sons, Inc. Available from University Microfilms, Ann Arbor, Michigan.

SUDDEN

EXTREME

UNEXPECTED

PREDICTABLE

PSYCHOLOGICAL

RARE

Figure 1. Characteristics of Catastrophes

SUDDEN

Donora vs. Los Angeles Smog

Iran Earthquake vs. Erosion

Epidemics vs. Normal Death Rate

Air Crash vs. Automobile Deaths

Poliomyelitis vs. Cardiac Arrest

Figure 2. Illustrative Examples

EXTREME

Date	Place	Type	Deaths
1947	Texas City	Fire-exp.	561
1942	Boston	Club Fire	492
1944	Port Chicago	Ammo Ships	322
1944	Hartford	Circus Fire	168
1947	Southwest	Tornado	167

Figure 3. Illustrative examples

UNEXPECTED

Thalidomide Teratogency

Pearl Harbor Attack

Prince of Wales and Repulse

Virginia Flood (1969)

Figure 4. Illustrative examples

PREDICTABLE

Radium illuminated dials

Giant Solar Flare

Unknown Moon Agent

Smallpox among natives

Nuclear test triggered tsunami

Figure 5. Illustrative examples

PSYCHOLOGICAL

Battle of Big Bethel

Sheep deaths near Dugway

D. C. Subway deaths

"Brainwash" political death

Figure 6. Illustrative examples

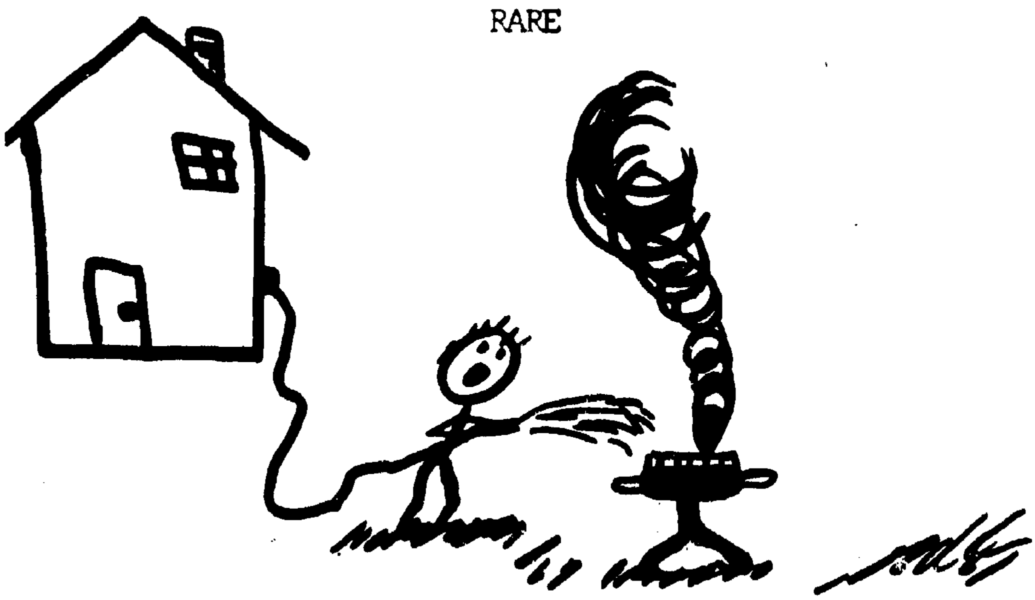


Figure 7. Putting out fires

RANGE	- - - - - - - - - - - - - -	all
TECHNIQUE	- - - - - - - - - - - -	"educated" guess
SAMPLE SIZE	- - - - - - - - - - -	unknown
VALIDATION-	- - - - - - - - - - -	experience

Figure 8. Common sense

Preceding page blank

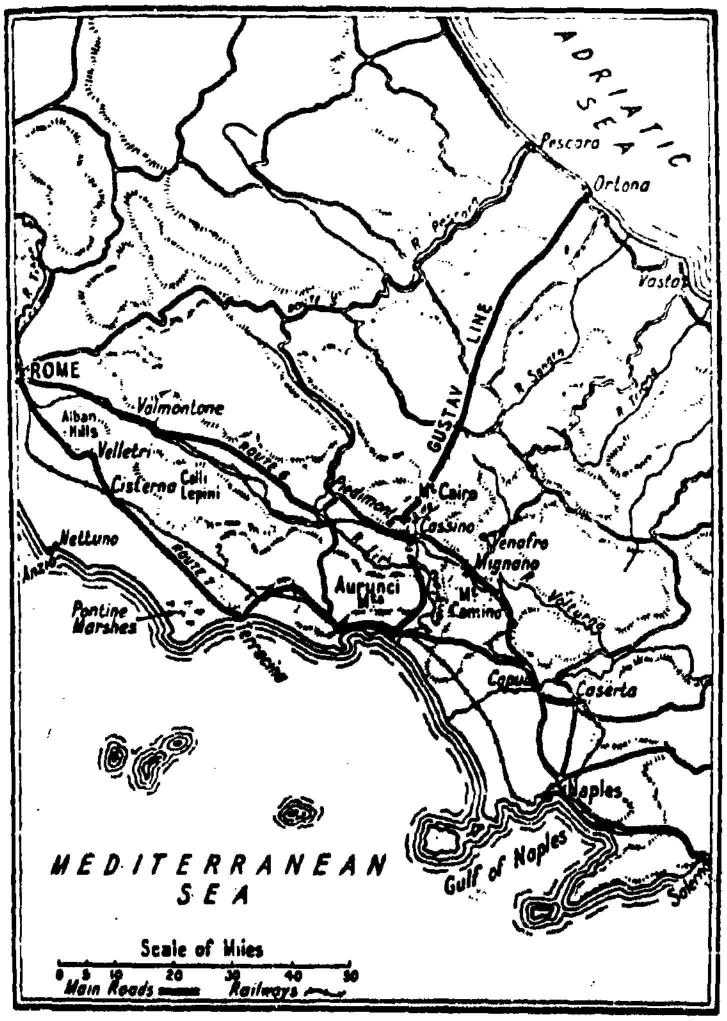


Figure 9. Theater of Anzio campaign

RANGE - - - - - - - - middle

TECHNIQUE - - - - - formal statistics

SAMPLE SIZE - - - - large

VALIDATION- - - - - experiment

Figure 10. Formal statistics

RANGE - - - - - - - - middle

TECHNIQUE - - - - - quality control

SAMPLE SIZE - - - - large

VALIDATION- - - - - feed back

Figure 11. Process monitoring

RANGE - - - - - - - - middle

TECHNIQUE - - - - - forecast

SAMPLE SIZE - - - - none

VALIDATION- - - - - feed back

Figure 12. Technological forecasting

RANGE - - - - - - - - extreme

TECHNIQUE - - - - - - - - -

SAMPLE SIZE - - - - none

VALIDATION- - - - - feedback

Figure 13. Catastrophic threat

PROBABILITY FOUNDATIONS

Necessary	Fermat, Pascal
Frequency	Graunt, von Mises
Abstract	Kolmogoroff
Personal	Bernoulli, Bayes
Neglected Causes	

Figure 14.

Cond	Rslt	Cond	Rslt	Cond	Rslt
A	F	B	S	C	S
A	S	B	S	C	F
A	F	B	S	C	F
A	F	B	F	C	S
A	S	B	S	C	F
A	F	B	S	C	F

Figure 15. Examples of chance sequences

$$\begin{matrix} P_{11} & \cdots & P_{12} & \cdots & P_{1j} & \cdots & P_{1c} \\ P_{21} & \cdots & P_{22} & \cdots & P_{2j} & \cdots & P_{2c} \\ . & \cdots & . & \cdots & . & \cdots & . \\ P_{r1} & \cdots & P_{r2} & \cdots & P_{rj} & \cdots & P_{rc} \end{matrix}$$

Figure 16. Matrix of probabilities assigned to catastrophes (columns) by different procedures (rows)

$$
\begin{array}{llll}
K_{11},\ P_{11} & K_{12},\ P_{12} & \cdots\cdot & K_{1c},\ P_{1c} \\
K_{21},\ P_{21} & K_{22},\ P_{22} & \cdots\cdot & K_{2c},\ P_{2c} \\
\cdot\quad\cdot & \cdot\quad\cdot & \cdots\cdot & \cdot\quad\cdot \\
K_{r1},\ P_{r1} & K_{r2},\ P_{r2} & & K_{rc},\ P_{rc}
\end{array}
$$

Figure 17. Matrix of catastrophes assigned a common probability

KOEHLER EFFECT

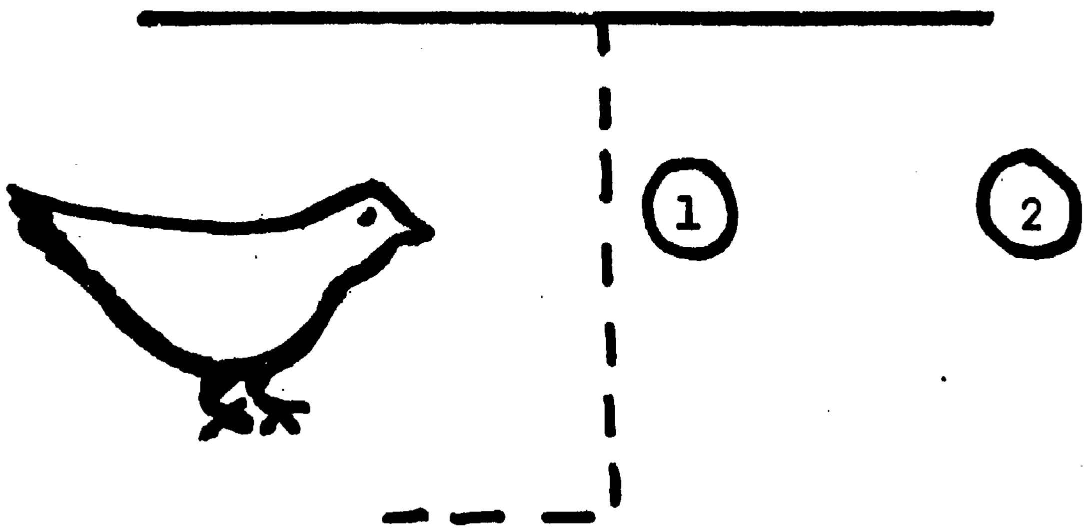


Figure 18. Effect of emotional intensity on decision-making

Feedback necessarily occurs

Information feedback is cheapest

Information feedback is ignorable

Feedback through channels is no feedback

Figure 19. Principles of decision-making verification

APPENDIX

1. A commander can make a decision simply by ruling out what appears to him to be impractical or unfeasible. (244)

2. General Alexander felt that ... allied troops on the enemy flank below Rome might so threaten German communications as to compel the enemy to retreat. (248)

3. Link-up between the main and the Anzio fronts, it was assumed, would take place no later than seven days after the landing. (248)

4. A strengthened Anzio force, if assured continuous resupply by water, could, he believed, consolidate a beachhead.... (249)

5. Whether the 60 miles between Anzio and the Garigliano was too great a distance for action on one front to influence the other was discussed, but it was accepted as an unavoidable risk. (249)

6. It was, impossible to predict the exact German reaction to a landing, but the most probable reactions seemed desirable from the Allied standpoint. (250)

7. The Anzio force might provoke the Germans ... to withdraw. (250)

8. ... intelligence officers of the 15th Army Group were rather optimistic. (251)

9. ... they 'counted on' the effect of weather and on harassment by the Allied air forces to interfere.... (251)

10. The [ambiguity of orders to Sixth Corps] arose from the difficulty of judging .... (252)

11. Fifth Army intelligence estimates were less optimistic.... (252)

12. The enemy was judged to have. (252)

13. By the third day the Germans could perhaps. (252)

14. Two additional divisions could probably. (252)

15. The Fifth Army assumed that the VI Corps would meet strong resistance on the beaches (252)

16. It expected the Corps to receive heavy counterattacks.... (252)

17. ... having underestimated German strength at Salerno.... (252)

18. The Fifth Army - and with it the VI Corps - expected the same pattern.... (252)

19. The Fifth Army expected the VI Corps to be ready to do one of two things upon landing. (253)

20. The operation becomes such a desperate undertaking. (254)

21. Otherwise "a crack on the chin is certain" (254)

22. A failure now would ruin Clark, probably kill me, and certainly prolong the war... (254)

23. A week of fine weather at the proper time and I (Lucas) will make it.... (255)

24. Alexander told Lucas "we have every Confidence in you" (255)

25. What troubled General Lucas...was the contrast between his own concern...and nonchalance in the higher echelons.... (255)

26. Lucas was not so sure. (255)

27. The chances are seventy to thirty that (256)

28. He [Lucas] believed his forces lacked the strength (256)

29. The general idea seems to be... (256)

30. I wish the higher headquarters were not so optimistic.... (256)

31. Securing a beachhead was all Fifth Army expected.... (257)

32. Lucas [was not] to push on to the Alban Hill mass at the risk of sacrificing his corps (257)

33. Such a possibility [moving on the Alban Hills] appeared slim to the Fifth Army Staff (257)

34. The staff questioned Lucas' ability.... (257)

35. It was obvious what the loss of the supply base would mean (257)

36. If the enemy came to Anzio in strength (257)

37. The British feared they might mistake Americans for Germans (257)

38. What everyone had overlooked ... was the possibility of achieving complete surprise (258)

39. The Germans always regarded the long sea flanks in Italy as exposed.... (258)

40. To reinforce [local troops Kesselring] expected to call on Tenth Army for a division.... (259)

41. He hoped to have the Fourteenth Army in north Italy move ... the equivalent of about one or two divisions (259)

42. Fearing that the Fifth Army was about to make a breakthrough.... (259)

43. Feeling that the fate of the Tenth Army .... (259)

44. According to the German estimate the landing had a good chance... (259)

45. Field Marshall Kesselring assumed that the troops would probably try to seize the Alban Hills (260)

46. The Germans were considerably reassured by Allied behavior at the landing (261)

47. Kesselring's order to stand fast on the Garigliano-Rapido line was... in the nature of a gamble.... (261)

48. If the Allies attacked on January 23 or 24, German forces would not be strong enough to hold (261)

49. The evening of 23 January, Kesselring "believed" that the danger of a beachhead expansion was no longer imminent (261)

50. By 24 January the German command considered the danger of an Allied breakthrough removed. (261)

51. Alexander was very optimistic, Clark somewhat subdued (261)

52. Lucas' concern with logistical aspects came not only from prudence (262)

53. He believed the Germans could increase their build-up (262)

54. He believed the Germans would stop his VI Corps before it could cut their line of communication (262)

55. His intelligence officers informed him that the Germans were taking troops from the Fifth Army main front to oppose him (262)

56. This might permit the Fifth Army to advance (262)

57. The Fifth Army, Lucas was certain, would still have to fight powerful rear guards (262)

58. He expected no spectacular rapidity of movement (262)

59. ... he sought to build up his strength and his supplies to remain intact even though isolated (263)

60. I feel now [January 25] the beachhead is safe (263)

61. Lucas expected the 1st Armored Division to arrive soon (263)

62. That is about all I can supply but I think it will be enough (263)

63. I must do nothing foolish (263)

64. I must hold it "..." I think I can (263)

65. Kesselring came to the conclusion that the Allies were preparing a full scale attack (263)

66. The best defense, he felt, was an attack on his own (263)

67. Lucas thought he could attack in a few days (264)

68. He expected 30 LST's to be unloaded at Anzio 27 January (264)

69. Clark "received the impression" that the outcome of the struggle depended on who could increase his forces more quickly (264)

70. Though the situation was not clear to Clark (264)

72. Apparently, some of the higher levels think I have not advanced with maximum speed (264)

73. I think more has been accomplished than anyone had a right to expect. (264)

71. He urged Lucas to take bold offensive action (264)

74. This venture was always a desperate one (264)

75. I could never see much chance for it to succeed (264)

76. Without Anzio our situation would have been desperate (264)

77. Had I rushed troops to Albano and Velletri they would have been destroyed (264)

78. The only thing to do was what I did (264)

79. Keep the enemy off balance until the Corp was ashore and everything was set (264)

. Lucas could envisage holding some troops in corps reserve (265)

. The situation is crowded with doubt and uncertainty (265)

. I expect to be counterattacked in the morning. (265)

. I think he realized the seriousness (265)

33. He [Clark] thinks I should have been more agressive on D-Day (265)

. There has been no chance to build "Shingle" up to decisive strength (265)

. Anyone could have seen that from the start (265)

. I can win if I am left alone (265)

. I don't know whether I can stand the strain (265)

. Clark and those above him thought Anzio would shake the Cassino [illegible] at once (266)

. They had no right to think that (266)

. It was clear that the attack had not accomplished much (266)

. The Ranger force met unanticipated opposition (266)

. The enemy had an unexpectedly strong and well organized defensive [illegible] (266)

. It seemed clear the Germans has built up their forces around Anzio (266)

. What the Allies did not know was how close they came to breaking out [illegible] beachhead (266)

. [illegible] Allied intelligence officers seemed like overwhelming German [illegible]

. Allied intelligence officers had to assume (266)

. Lucas thought he could support two more divisions at Anzio on [illegible] (267)

. [illegible] and Clark decided that the enemy build-up dictated a [illegible] defensive tactics (267)

. [illegible] February "keeping the enemy off balance" was a forlorn hope (267)

. [illegible] are disappointed but there is no military reason why they [illegible]

102. General Devers thought Lucas should have gone on - on landing (268)

103. "Had I done so, I would have lost my corps" (268)

104. Clark thought Lucas had done all he could at Anzio (268)

105. I thought I was winning something of a victory (268)

106. General Clark thought Lucas could have taken the Alban Hills but could not have held them (268)

107. Clark thought British G-2 intelligence was always over optimistic (269)

108. The Germans built up their differences at Anzio much faster than the British believed possible (269)

109. Clark had always felt that Anzio had little chance of success (269)

110. In retrospect Clark felt that the total losses at Garigliano and at Anzio might have been safer and as productive at Garigliano alone (269)

111. A powerful counter attack at Anzio could well have wrecked the entire Italian Campaign (269)

112. By the end of January, Clark was disappointed by Lucas' lack of aggressiveness (269)

113. Clark believed Lucas should have made a reconnaissance in force to capture Cisterna and Campoleone.... (269)

114. Clark thought such an effort to be not incommensurate with Lucas' forces (269)

115. Others felt much the same (269)

116. General Marshall thought Lucas could have taken the Albán Hills (269)

117. However, he thought Lucas had acted wisely (269)

118. Marshall felt Lucas could not have held the Alban Hills and the port at Anzio (269)

119. The theater G-2 had held the same opinion at the Christmas Day Conference at Tunis (269)

120. G-2 thought Lucas would have been in a bad way without a main front breakthrough. (269)

121. The Allies would be unable to keep the Germans from shifting forces to Anzio from south Italy as well as elsewhere (269)

122. General Lemnitzer also felt the Allies did not have the strength to hold the Alban Hills (269)

123. Lemnitzer thought that Alexander hoped that the Anzio operation plus a main force attack "might" force a German withdrawal (270)

124. The advance "on" the hills was exactly what Alexander thought possible (270)

125. When Alexander visited the beachhead on D-Day he approved the decision not to push out far from Anzio. (270)

126. Lemnitzer thought that Alexander thought that Lucas had done no wrong but was under too much strain (270)

127. By that time it was clear the Anzio operation would involve a long, hard struggle (270)

128. It would seem that Lucas' action during the first few days was justified (270)

129. The main German Army showed no signs of withdrawing (270)

130. The Allies saw no immediate prospect of forcing a general retreat (270)

131. It became far more likely the Germans would move in strength against Anzio (270)

132. If the VI Corps went too far inland it would risk annihilation (270)

133. Allied intelligence judged the German strength as sufficient but not overwhelming (270)

134. It would seem that the Allied hesitation on the Anzio shore stemmed from a belief in German invincibility (270)

135. This belief was a product of doubt and uncertainty both before and during the operation (270)

136. This belief was used later to explain the inevitability of the actual course of events (270)

137. The only thing that disturbed Lucas was the necessity to safeguard the port (271)

138. Without it the swift destruction of the corps was inevitable (271)

139. Lucas thought he could not have done differently (271)

140. Nevertheless the alternative remained a disturbing possibility to him (271)

141. He admitted a mass of armor and motorized infantry might have reached the Alban Hills (271)

142. He was sure he could not have remained there (271)

143. Any force that far from Anzio would have been in the greatest jeopardy (271)

144. Lucas did not see how it would have escaped annihilation (271)

145. As it turned out he believed he had reached positions from which the enemy was unable to dislodge him (271)

146. Lucas believed the whole operation a mistake (271)

147. Anyone who expected him to push to the Alban Hills was bound to be disappointed (271)

148. Lucas had never considered doing so (271)

149. He considered his mission to be taking the port and its surroundings (271)

150. Perhaps this was an influence of the Navy (271)

151. Admiral Cunningham asserted no reliance could be placed on over the beaches maintenance (271)

152. Unfavorable weather was probable (271)

153. General Clark said: "You can forget this goddam Rome Business" (271)

154. The capture of Anzio was an obvious objective (271)

155. But early occupation of the Alban Hills was vital (271)

156. The Anzio forces later realized the importance of the hills (272)

157. Was General Lucas justified in delaying seven days before starting his offensive? (272)

158. Could he have gotten away with the gamble of an immediate drive to the Alban Hills (272)

159. Certainly the complete surprise achieved at the landing could have been exploited (272)

160. According to Tenth Army estimates only a quick cutting of lines of communication would have led to major Allied success (272)

161. Such a success would be more likely to capture Rome (272)

162. According to Kesselring's Chief of Staff, an audacious flying column could have penetrated to the city (272)

163. He was astonished at the Allied passivity (272)

164. Could the Germans have withstood a dynamic front as they did the static front? (272)

165. Would they have dared to hold both at Anzio and at Garigliano (272)

166. An Allied force ensconced on the Alban Hills would have been a much greater threat than those or Anzio (272)

167. The answer can only be speculation (272)

168. Alexander thought an aggressive commander would have acted differently than Lucas (272)

169. He would and could have pushed regimental strength patrols to the hills (272)

170. The shock of Allied troops directly threatening Rome might have by itself permitted Allied retention of both the hills and a supply corridor (272)

171. A bluff might have worked (272)

172. General Patton might have been successful (272)

FROM: COMMAND DECISIONS - Office of the Chief of Military History
Department of the Army 1959

EMPIRICAL BAYES AND THE DESIGN AND
ANALYSIS OF EXPERIMENTS

Richard G. Krutchkoff
Virginia Polytechnic Institute
Blacksburg, Virginia

Let me start with an introduction to Empirical Bayes. Consider the simple estimation situation in which we observe a value x of the random variable X which has distribution function $F(x|\theta)$, and must estimate θ with small squared error. In the parametric situation, the form of the distribution function is known except for the value of the parameter θ. Both x and θ may be vector valued.

For Empirical Bayes to be applicable here we consider the case in which the estimation problem is routine. That is, we observe x_1 from $F(x_1|\theta_1)$ and must estimate θ_1; then some time later, in a similar but independent situation, we observe x_2 from $F(x_2|\theta_2)$ and must estimate θ_2. This routine situation continues until at present we have the observation x_n from $F(x_n|\theta_n)$ and we must estimate θ_n. These estimating situations we call experiences. As an example consider the situation encountered at the Radford Arsenal. Every six weeks base grain was mixed and subsequently cured with Nitroglycerine in order to form propellent for the Nike missile. It was desired to estimate the parameters for each base grain separately since it was believed that the parameters would vary from base grain to base grain in some unpredictable manner. Since the θ values θ_1, θ_2, $\theta_3, \ldots, \theta_n$, vary in an arbitrary and unpredictable manner, we assume that θ is a random variable but with a completely unknown distribution. It is important to note that we do not use our ignorance of the distribution as a justification for choosing a diffuse or uniform distribution.

If one were to take a completely classical approach to the problem he would note that X_n is a sufficient statistic for θ_n. This is easily seen by noting that

$$F(x_1,x_2,x_3,\ldots,x_n|\theta_1,\theta_2,\theta_3,\ldots,\theta_n) = \prod_{i=1}^{n} F(x_i|\theta_i) \quad .$$

The classical solution to the problem therefore must completely ignore the observations $x_1,x_2,x_3,\ldots,x_{n-1}$ and use only the observation x_n in estimating θ_n. Even intuitively this is an unfortunate result.

The pure Bayesian approach to the problem assumes a form for the distribution of θ, say $G(\theta)$, and then obtains the estimator

$$E(\theta|x_n) = \frac{\int \theta dF(x_n|\theta)dG(\theta)}{\int dF(x_n|\theta)dG(\theta)}$$

(the posterior mean) as the minimizing estimator. If the choice of $G(\theta)$ is correct then this is indeed the minimizing estimator. If the choice of $G(\theta)$ is not correct then the estimator may have a very large mean squared error. Note that the Bayes estimator ignores the past experience $x_1, x_2, x_3, \ldots, x_{n-1}$, as did the classical estimator. Surely, we should be able to use this experience in some way.

The Empirical Bayes approach to this problem is now very simply stated. We find the Bayes estimator $E(\theta|x_n)$, which is usually given in terms of the unknown distribution function, and express it in a form which can be estimated from the data, $x_1, x_2, x_3, \ldots, x_n$ without knowledge of, or assumptions about, the unknown prior distribution. The proper forms for $E(\theta|x_n)$ are given in Rutherford and Krutchkoff [11] for four general families of distributions. Examples of members of these general families are the Poisson, Negative Binomial, Logarithmic, Gamma, Normal (unknown mean), Normal (unknown variance), Exponential, and the Uniform Distributions. In Lemon and Krutchkoff [5] an Empirical Bayes estimating procedure is proposed for any discrete conditional distribution. This procedure has now been extended to include any conditional distribution.

Let me now briefly mention some recent applications of this approach. First, consider the simple linear orthogonal model

$$Y_i = \alpha + \beta(X_i - \bar{X}) + \varepsilon_i$$

where the errors are assumed to be normal and where we must routinely estimate α and β. This problem is considered in Clemmer and Krutchkoff [1] and the example analyzed there is worth rementioning here.

Every six weeks Radford Army Arsenal mixed Base Grain for their Nike missiles. The Base Grain was then cured with Nitroglycerine to form rocket propellent. Estimates of the parameters in a linear model were required for each Base Grain. Since the chemicals were purchased at different times and mixed at different times under different atmospheric conditions, the parameters were expected to vary in an unpredictable manner. Using the estimator for the normal distribution given in Rutherford and Krutchkoff [11], Clemmer and Krutchkoff [1] found the desired estimators as

$$E(\alpha|\hat{\alpha}) = \hat{\alpha} + \frac{\sigma^2}{N}\frac{f'(\hat{\alpha})}{f(\hat{\alpha})}$$

and

$$E(\beta|\hat{\beta}) = \hat{\beta} + \frac{\sigma^2}{S_{XX}}\frac{f'(\hat{\beta})}{f(\beta)}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the usual Least Squares or Maximum Likelihood estimators for α and β, σ^2 is the error variance, N is the number of observations taken in the n^{th} experience, S_{XX} is the usual sum of squares of the independent variable $(\sum_{i=1}^{n} (X_i-\bar{X})^2)$; $f(\hat{\alpha})$ is the marginal density of the least squares estimator; and, $f'(\hat{\alpha})$ is the derivative of this density estimated at the same point. In general, the form of the estimator is simply the least squares estimator plus a correction factor. The correction factor is the variance of the least squares estimator times the ratio of the derivative of the marginal density to the marginal density itself evaluated at the present value of the least squares estimator. It is worth noting here that if the parameter has a diffuse prior distribution then the ratio of the derivative of the density to the density will in effect be zero and the Empirical Bayes estimator will be the Classical estimator. Thus, when the prior information is of little value this correction term disappears rather than biasing the result undully.

The estimate of the ratio recommended in the paper [1], can be simply written as

$$\frac{f'(\hat{\alpha})}{f(\hat{\alpha})} = \frac{\sum_{i=1}^{n}\left\{\left[\frac{S_{in}A_i^*}{A_i^*}\right]^2 - \left[\frac{S_{in}A_i}{A_i}\right]^2\right\}}{h\sum_{i=1}^{n}\left[\frac{S_{in}A_i}{A_i}\right]^2}$$

where

$$A_i = \frac{\hat{\alpha}_n - \hat{\alpha}_i}{2h}$$

$$A_i^* = \frac{\hat{\alpha}_n - \hat{\alpha}_i + h}{2h}$$

and where

$$h = n^{-1/5}\left\{\max\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_i-\bar{\hat{\alpha}})^2,\ \frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2\right]\right\}$$

with $\bar{\hat{\alpha}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i$, and $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$.

Note that we are using $\hat{\alpha}_i$ to represent the least squares estimate for α in the i^{th} experience and $\frac{\sin 0}{0}$ is to be interpreted as unity. An estimate of $\frac{f'(\hat{\beta})}{f(\hat{\beta})}$ is obtained by simply replacing the α's with β's.

The improvement of this Empirical Bayes estimating procedure is then compared with the Classical procedure by taking a ratio of the mean squared errors. This was done by choosing a distribution for α and β and generating α_1, β_1 from this distribution and then generating several observed values y_i from the regression equation $y_i = \alpha_1 + \beta_1(x_i-\bar{x}) + \epsilon_i$ with ϵ_i being random normal errors. This was done fifty times obtaining new values for the observations at each experience. The Empirical Bayes estimator was obtained using the i-1 previous sets of data as past experience. The entire run of 50 experiences was then repeated 500 times. The average ratio of the Empirical Bayes squared errors to the Classical variance was then plotted as a function of the number of experiences. This was then repeated for many different prior distributions, error variances and experimental designs. It was found that the mean squared error for the Empirical Bayes procedure was never greater than that of the classical procedure with the ratio of the Empirical Bayes mean squared error to the Classical mean squared error often dropping well below unity. It was also determined that the ratio of the mean squared errors depends not on the prior distribution or the error variance or the design but solely on one relation involving them; namely,

$$Z = \frac{\mathrm{Var}(\hat{\alpha}|\alpha)}{\mathrm{Var}\ \alpha}$$

for α and a similar expression for β. This index is the ratio of the least squares variance to the variation in the parameter. Intuitively, if the least squares variance is small and the parameter variation large not much information can be extracted from past experience. This is, in fact, the case. When Z is extremely small, below 0.1, we find that the derivative of the marginal on α is small compared with the density itself and the correction factor disappears. On the other hand, when the least squares variance is large and the parameter variation small, much is to be gained from past experience. However, when Z is very large (say 10) then one might as well assume that the parameter is not varying at all and pool all the data. The interesting and realistic range is when Z is about unity. Figures 1, 2, and 3 given here are for Z values of 0.5, 1 and 2 with past experience ranging from one to fifty. The solid line is for σ^2 known, obtained by pooled data or estimated from the present data with $N \geq 20$. The broken line is for σ^2 estimated from the present set of data with $N = 8$. The reduction in mean squares error obtained from nineteen batches of Base Grain is given here in Table 1.

Then in Martz and Krutchkoff [6] the regression model was extended to the multilinear model

$$y_i = \alpha + \beta X_i + \gamma X_i^2 + \dots + \varepsilon_i$$

where orthogonality was not required. This required obtaining the multivariate extension to the estimators presented in Clemmer and Krutchkoff [1] and finding estimators for joint marginal densities and their vector derivatives. The mean squared errors once again were never greater than those of the least squares estimators with their ratio often dropping well below unity for the usual Z values.

The model was then extended to allow for the possibility that σ^2 varies in an unpredictable way from experience to experience. Consider the model

$$\underset{\sim}{y} = \underset{\approx}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

where $\underset{\approx}{X}$ is a $k \times p$ matrix of known fixed quantities which remains the same from experiment to experiment and $\underset{\sim}{\varepsilon}$ is distributed $N(\underset{\sim}{0}, \sigma^2\underset{\approx}{I})$. We assume $\underset{\sim}{\beta}$ and σ^2 vary randomly from experiment to experiment according to the unknown prior distribution $G(\underset{\sim}{\beta},\sigma^2)$.

If $\underset{\approx}{X}$ is of rank p, the usual least squares estimators for $\underset{\sim}{\beta}$ and σ^2 are

$$\hat{\underset{\sim}{\beta}} = (\underset{\approx}{X}'\underset{\approx}{X})^{-1}\underset{\approx}{X}'\underset{\sim}{y}$$

and

$$\hat{\sigma}^2 = \frac{(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})' (\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta})}{k-p}$$

Denote $(k-p)\hat{\sigma}^2$ by S. For this situation, the Empirical Bayes estimators are given by

$$\tilde{\underset{\sim}{\beta}} = \hat{\underset{\sim}{\beta}} + \frac{S(X'X)^{-1}}{k-p-2} \frac{f_{N,\hat{\beta}}\,, {}_{k-p-2}(\hat{\underset{\sim}{\beta}},S)}{f_{N,k-p}(\hat{\underset{\sim}{\beta}},S)}$$

and

$$\tilde{\sigma}^2 = \frac{S}{k-p-2} \frac{f_{N,k-p-2}(\hat{\underset{\sim}{\beta}},S)}{f_{N,k-p}(\hat{\underset{\sim}{\beta}},S)}$$

The past experience for this is in the form of the vectors

$$\begin{pmatrix} \hat{\underset{\sim}{\beta}}_1 \\ S_1 \end{pmatrix}, \begin{pmatrix} \hat{\underset{\sim}{\beta}}_2 \\ S_2 \end{pmatrix} \dots \begin{pmatrix} \hat{\underset{\sim}{\beta}}_n \\ S_n \end{pmatrix} .$$

The ratio of densities given here are estimated in a way similar to the expressions already given but a bit more complicated. The actual formulas are not yet published, but can be found in the Virginia Polytechnic Institute Ph.D. dissertation of one of my students (see Rencher [8]). Needless to say, many simulations were run and never was the Classical squared error for any component of $\underset{\sim}{\beta}$ smaller than that of the Empirical Bayes procedure for as few as one past experience. The amount of improvement was similar to the figures already shown. On the other hand, there were cases in which we needed as many as five experiences before the Empirical Bayes procedure had a smaller squared error than the Maximum Likelihood procedure when estimating σ^2. See for example figure 4.

Another example of an Empirical Bayes application is in Sequential Estimation. Here we considered the case in which one must sequentially estimate the mean of a Normal distribution, the cost being the sum of the mean squared error and a constant times the number of observations taken. Although the big problem was the stopping rule, we had some difficulty in handling the past experience, since the number of observations taken differed from experience to experience. This problem was solved, however, and the solution is generally applicable to this type of past experience. The results, i.e., the ratio of Empirical Bayes cost to Classical cost plotted as a function of the number of experiences is typically as shown in the solid line of figure 5. The dotted line is the improvement obtained by using the Classical stopping rule and then the Empirical Bayes estimator. Since determining the stopping time by the Empirical Bayes approach is so very tedious we recommend using this hybrid approach. Unfortunately, the details of these procedures have not as yet been submitted for publication. They are available, however, in the Ph.D. dissertation of another one of my students (see Lemon [4]).

Another project presently underway is the estimation of the power spectral density function in a time series. In a time series situation one often has past experience from similar situations or one can break the present time series into parts which can be considered experiences. For example, in testing stress on airplane wings in a wind tunnel one has the results of tests on other wings. When the Navy obtains a time series signal from the path of a submarine, it is merely one experience in many such experiences. In each of these the object is to obtain the power amplitude of the various frequency components. We have used the Empirical Bayes approach to obtain efficiencies of the order of 150% that of the standard approach. This work is still in progress.

A project which is just about complete now involves estimating the arrival and service parameters in a Que. We have estimators for Ques involving the exponential and the Ehrlang distributions. As usual, the Empirical Bayes estimators have a significantly smaller mean squared error than the usual estimators. A typical example is depicted in Figure 6. λ is the mean arrival time, μ the mean service time, and ρ is the traffic intensity for an M/M/1 Que.

Now let us discuss the Analysis of Variance. First, we considered the random effects model:

$$y_{ij} = \mu + a_i + \epsilon_{ij}$$

with I effects and J repetitions per effect. We were able to estimate the variance σ_A^2 of the effects as well as the error variance by using past experience. By making a ratio of these two statistics, we came up with an analog to the F statistic. The percentage points of this statistic were found by Monte Carlo simulation for several values of I, J and numbers of experiences. We found that these tables depend only on J and N and not on I. Many typical situations were then simulated and the power for the Empirical Bayes test of the Hypothesis $\sigma_A = 0$ was always significantly greater than that of the usual F-test for the same size. After as few as ten or fifteen past experiences the power was as much as 50-80% higher for the Empirical Bayes test.

For the fixed effects model

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

we had to reparameterize to full rank before proceeding. Once this was done, the estimates were the same as for the linear regression situation. In order to test the Hypothesis $\beta = 0$, we made an analogy to the F statistic by using the sum of squares of the Empirical Bayes estimators for the parameters in the numerator and the estimated error variance in the denominator. Here the percentage points were found to depend on the number of repetitions, the number of experiences and also the number of effects. Only tables for up to six effects were simulated. Once again the Empirical Bayes test was always more powerful than the usual F-test (when there were at least four past experiences). Unfortunately, the details of this topic are not yet in print, but can be found in Rencher [8].

Before leaving parametric Empirical Bayes, I've been asked to briefly mention the results we obtained in long range prediction of rainfall. The Weather Bureau puts out a map, twice a month, predicting rainfall in the categories of light, moderate and heavy for a period of 30 days. The predictions are for large areas and not for particular locations. We were asked to use this map and predict for each city the amount of rainfall within the next 30 days. We were able to do just this. We found a procedure for predicting the probability distribution of rainfall in inches for any location that had been collecting such data for at least fifteen years. The results were remarkably successful. One could only compare with the Weather Bureau, however, for the categories light, moderate, and heavy. The Weather Bureau, for example, was correct in Roanoke, Virginia, but 30% of the time while we were correct more like 70% of the time. The details of this project can be found in Philpot and Krutchkoff [7].

Let us now turn to another type of Empirical Bayes Estimation. Consider the situation where the distribution of the observation is itself unknown; a non-parametric situation. Here, we observe the value x of the random variable X whose distribution depends in some unknown way on θ and we are asked to estimate θ with small squared error. Such estimation is of course impossible. For this situation, we assume that there is a supplementary observation obtained after the estimate is given, perhaps in the form of customer feedback. To be more specific, let us say we have observed an x from some unknown distribution and must estimate the θ value related to it in some unknown way. Later, we are given an observation y from the random variable Y whose distribution may likewise be unknown but for which $EY = \theta$. That is, our supplementary sample is an unbiased estimate of θ. Unfortunately, this estimate is too late. In Krutchkoff [3] the problem is assumed to be routine with the θ values varying in an unpredictable way. The Empirical Bayes estimating procedure for this situation is very simple. If the present observed values is x and there are several past experiences with this same value of x, then use as your estimate the average of the supplementary values y which occurred after the occurrence of the value x. If there is not a sufficient number of past experience at $x_i = x$, then make a linear regression using the past y values as the dependent variable and the past x values as the independent variable and find the regression value of y at x. The results of such a situation are given in Krutchkoff [3]. Generally, after a few past experiences the Empirical Bayes mean squared error drops below the mean squared error of the Classical estimator which would be used if the distribution of X were actually known. Here we have not only an estimator which we can use when nothing else exists, but one which is better than the usual estimator when one does exist.

An extension of this non-parametric approach was given in Gabbert and Krutchkoff [2]. Here we assumed that a machine producing items was to be checked to determine when it was Out of Control. The linear regression form of the estimate was employed but using only the past fifteen experiences.

A sample of defectives was taken and x, the sample proportion of defectives found. The value y was later supplied by some other procedure such that $EY = p$, the true proportion of defectives. Clearly, each box is an experience with the true proportion of defectives varying randomly. The estimate of the present proportion was obtained from x by using the past fifteen values of x,y, in a linear regression, and obtaining the regression value of y for the present value of x. The variance needed in the control chart was also obtained from the linear regression as the variance of the regression line at the present value of x. A typical power function for the Empirical Bayes procedure is given in figure 7. Here the machine was caused to go out of control, producing a proportion of defectives varying randomly about some undesirable proportion P_1 (P_0 is the in control value). The power for the procedure is seen to start out below the usual control chart procedure but after a few experiences with the out of control machine the power increases rapidly. In this example, the Empirical Bayes procedure detects the out of control situation at about the twelfth consecutive sample whereas it takes the usual procedure about 18 samples. Of course, since this Empirical Bayes procedure does not make use of the fact that the sample is Binomial, it can be applied in situations where the distribution of the first sample is itself unknown. We are presently working on a single sample non-parametric approach but not enough results are as yet available to present anything here.

Although the title of this paper contains the "design of experiments." I have virtually nothing as yet to report. Several results, however, can easily be predicted. For the non-parametric supplementary sample situation we did not require the present supplementary sample to determine the estimator or its squared error. If the squared error is within tolerable limits, we need not take the supplementary sample at all. We can, in effect, calibrate the preliminary sample eliminating the need for talking the more costly supplementary samples.

In the parametric situation we recall that our squared error was smaller than that of the Classical procedure. By estimating the prior variance we can estimate the efficiency of this procedure and thus be able to predict the number of observations one needs to obtain a predetermined squared error. This number will, of course, be smaller than that required by the Classical Procedure.

No doubt there are optimal designs for the Empirical Bayes procedures. Since Empirical Bayes is more efficient than the Classical approach using the classically optimal design it makes good sense to hope for an even better efficiency when we find the Empirical Bayes optimal design. This question is, as yet, unanswered, but we are working on it.

REFERENCES

1. Clemmer, B. A. and Krutchkoff, R. G. (1968). The use of Empirical Bayes estimators in a linear regression model, BIOMETRIKA, Vol. 55, pp. 525-534.

2. Gabbert, J. T. and Krutchkoff, R. G. (1970). Supplementary Sample Non-parametric Empirical Bayes in a Quality Control Situation, BIOMETRIKA, to appear.

3. Krutchkoff, R. G. (1967). A supplementary Sample non-parametric Empirical Bayes approach to some statistical decision problems, BIOMETRIKA, Vol. 54, pp. 451-458.

4. Lemon, Glen H. "Emprical Bayes Analysis of Some Sequential Experiments." Virginia Polytechnic Institute Ph.D. Dissertation, December, 1968.

5. Lemon, G. H. and Krutchkoff, R. G. (1969). "An Empirical Bayes Smoothing Technique." BIOMETRIKA, Vol. 56, pp. 361-365.

6. Martz, H. F. and Krutchkoff, R. G. (1969). Empirical Bayes estimators in a multiple Linear regression model. BIOMETRIKA, Vol. 56, pp. 367-374.

7. Philpot, J. W. and Krutchkoff, R. G. (1969). Probability Forecasts of 30-Day Precipitation. Published by the Water Resources Research Center of Virginia, Bulletin #2, 145 pages.

8. Rencher, Alvin C. "The Empirical Bayes Approach to the Analysis of Variance and Linear Regression." Virginia Polytechnic Institute Ph.D. dissertation, August, 1968.

9. Rutherford, J. R. and Krutchkoff, R. G. (1967). The Empirical Bayes approach: Estimating the prior distribution, BIOMETRIKA, Vol. 54, pp. 326-328.

10. Rutherford, J. R. and Krutchkoff, R. G. (1969). ε asymptotic optimality of Empirical Bayes estimators, BIOMETRIKA, Vol. 56, pp. 220-222.

11. Rutherford, J. R. and Krutchkoff, R. G. (1969). Some Empirical Bayes techniques in point Estimation, BIOMETRIKA, Vol. 56, pp. 133-137.

TABLE 1

NIKE MISSILE RESULTS

MAXIMUM PRESSURE VS. AGE:

	FOR α:		FOR β:	
n	Z	R	Z	R
2	.43	.98	1.34	.93
3	.51	.95	1.72	.88
4	.45	.95	1.08	.87
5	.23	.97	.53	.90
6	.45	.91	.89	.85
7	.42	.91	.98	.83
8	.74	.86	1.38	.75
9	.64	.88	1.30	.75
10	.59	.86	1.23	.72
11	1.04	.77	5.01	.58
12	.95	.77	3.69	.56
13	.30	.91	.71	.81
14	.69	.81	1.72	.80
15	.54	.86	1.36	.70
16	.54	.86	1.57	.69
17	.41	.89	2.51	.61
18	.52	.86	1.57	.69
19	.17	.96	.45	.87

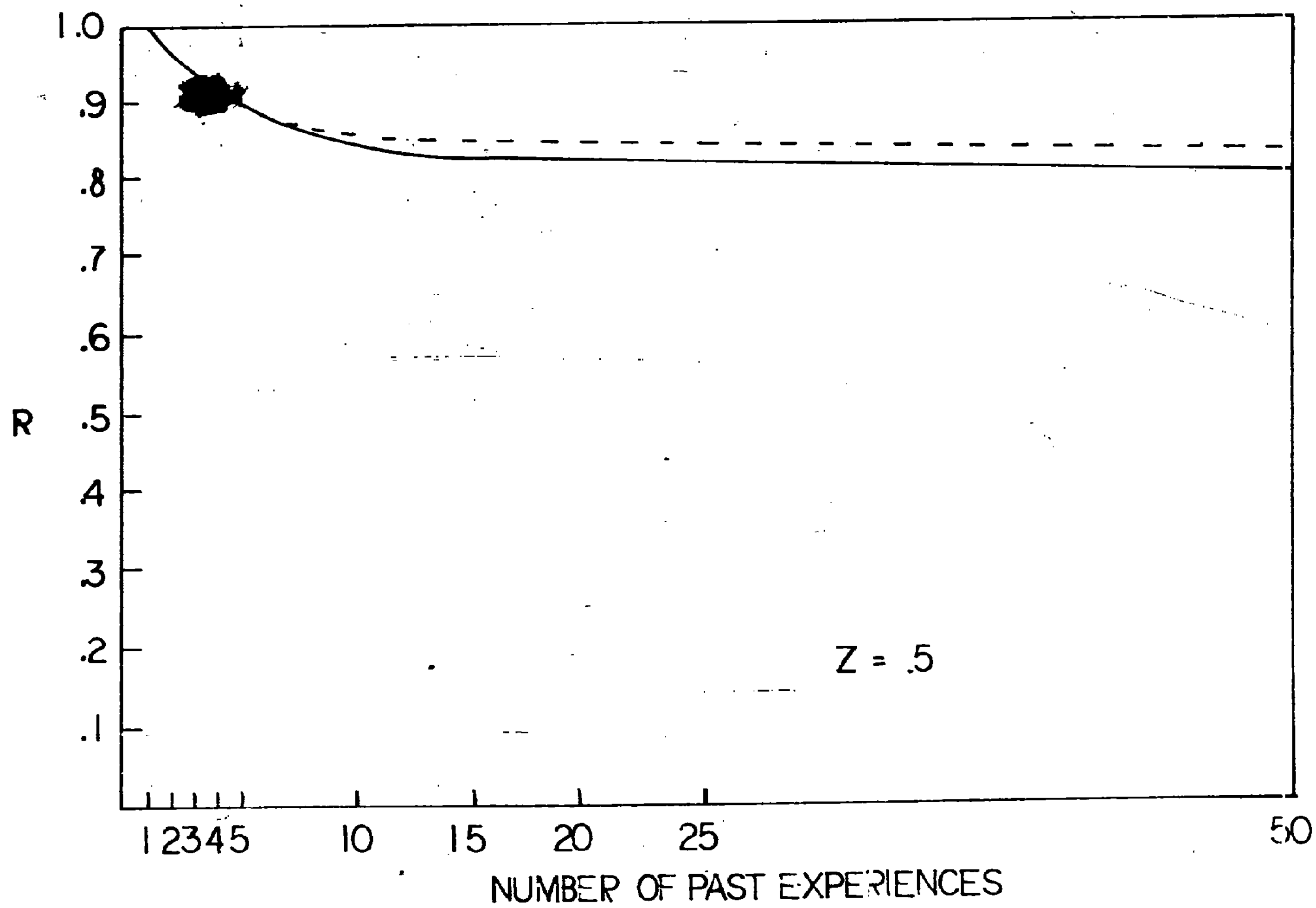


FIGURE I

Ratio of mean squared error of the Empirical Bayes estimator to the MSE of the least squares estimator for Z = .5

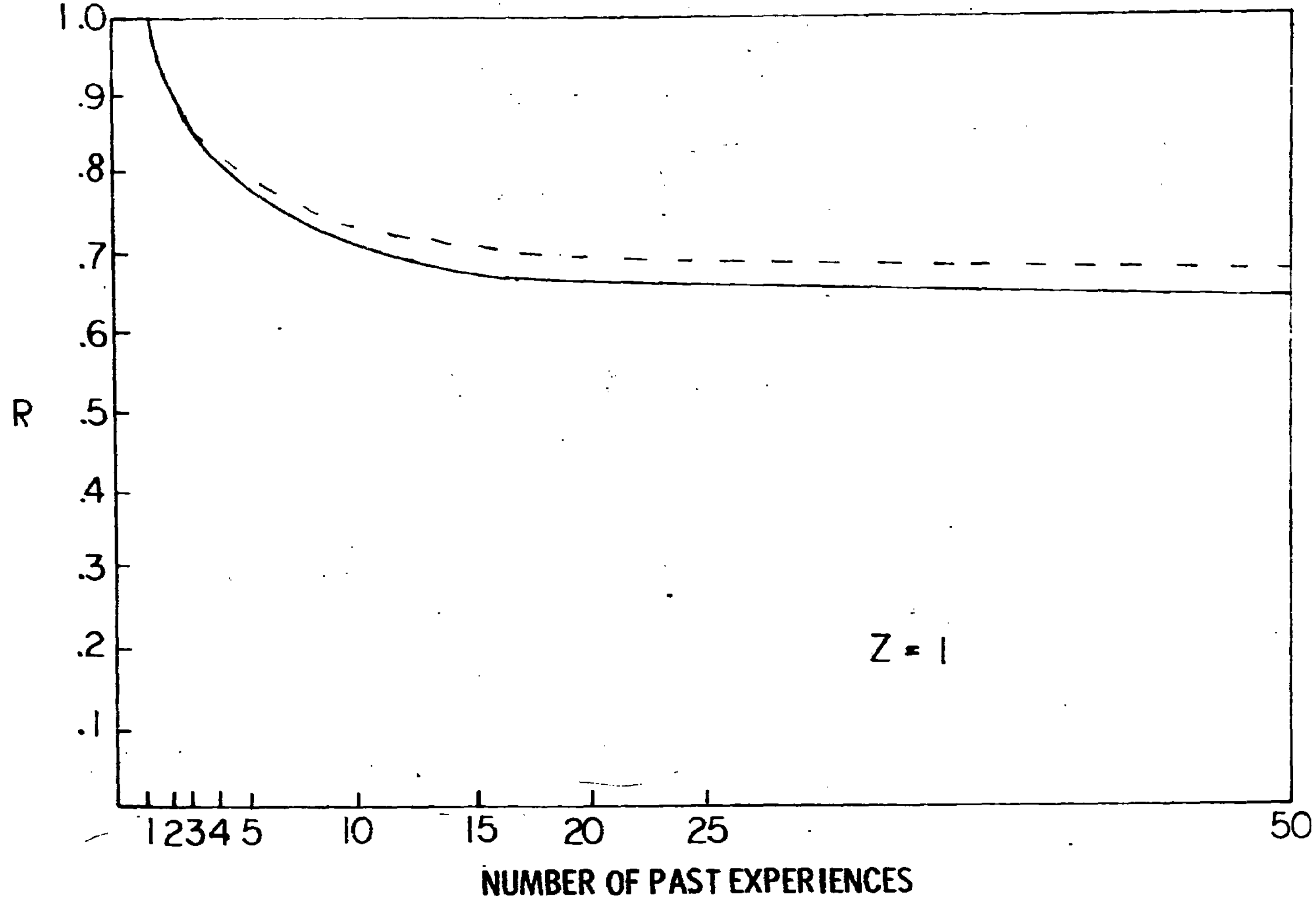


Figure 2
Ratio of the mean squared error of the Empirical Bayes estimator to the mean squared error of the least squares estimator for Z = 1

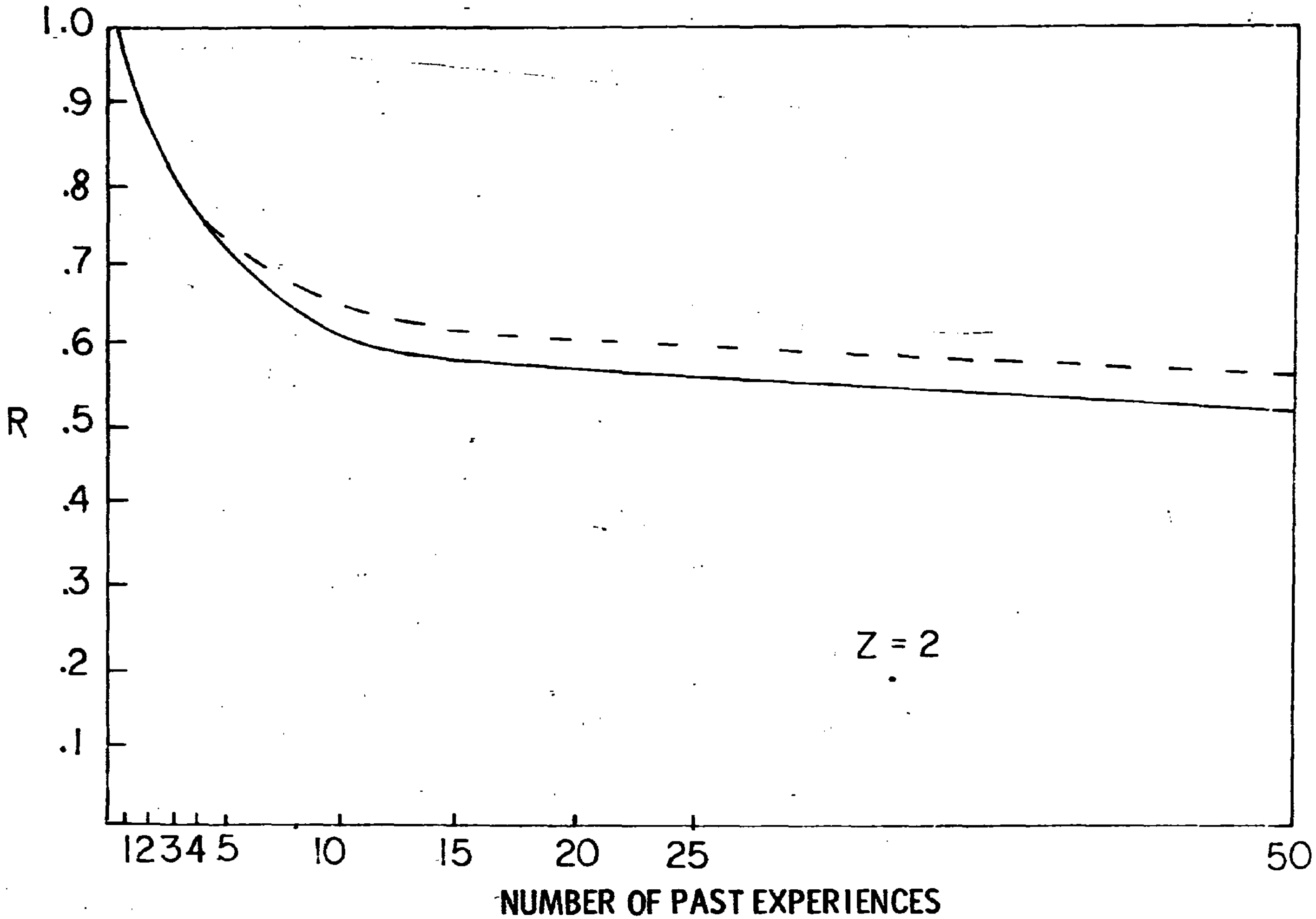


Figure 3

Ratio of the mean squared error of the Empirical Bayes estimator to the mean squared error of the least squared estimator for Z = 2

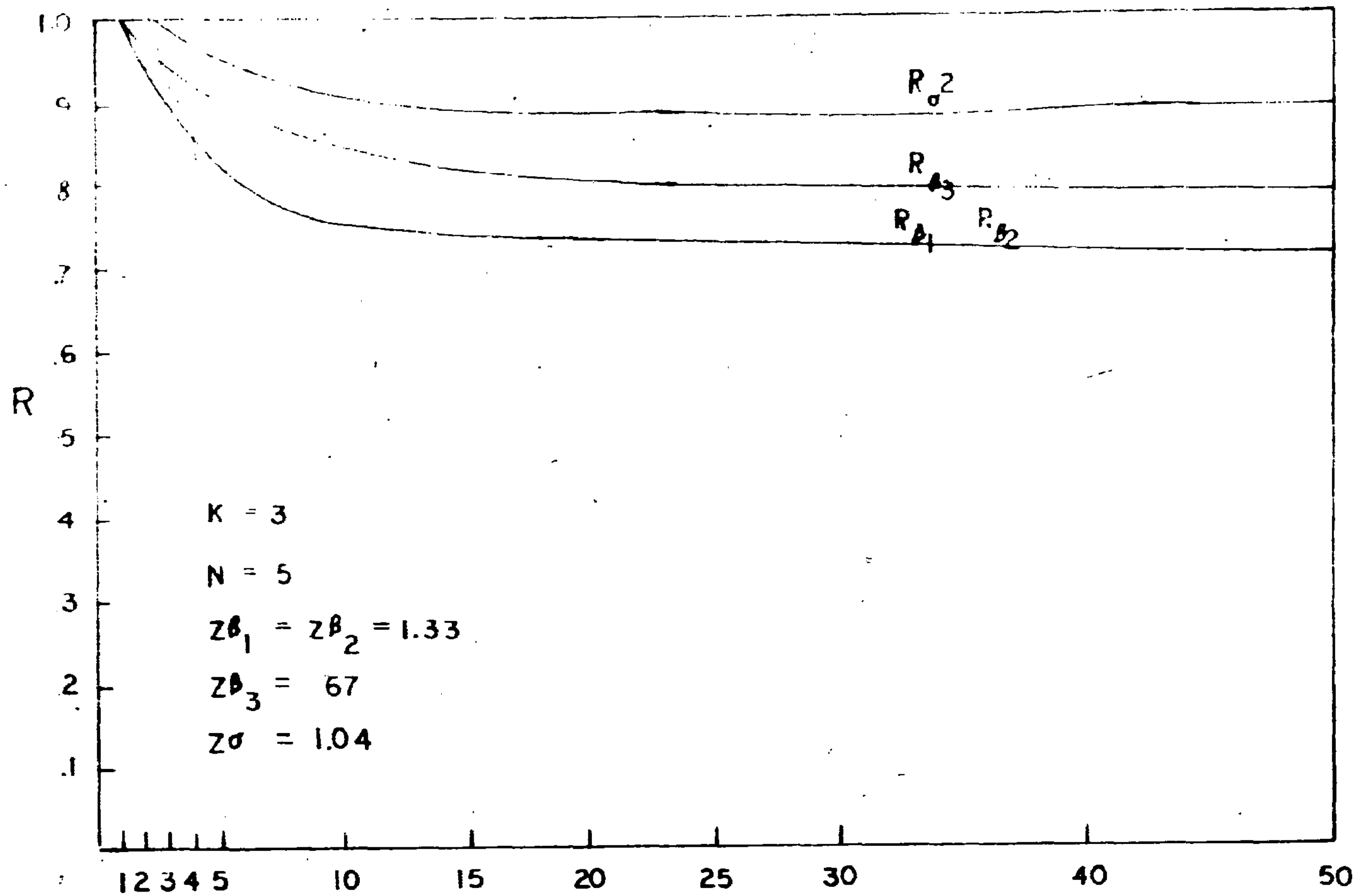


Figure 4

Ratio of mean squared error of the Empirical Bayes Estimators to the MSE of the maximum likelihood estimators for $_1$, $_2$, $_3$ and ^{2}

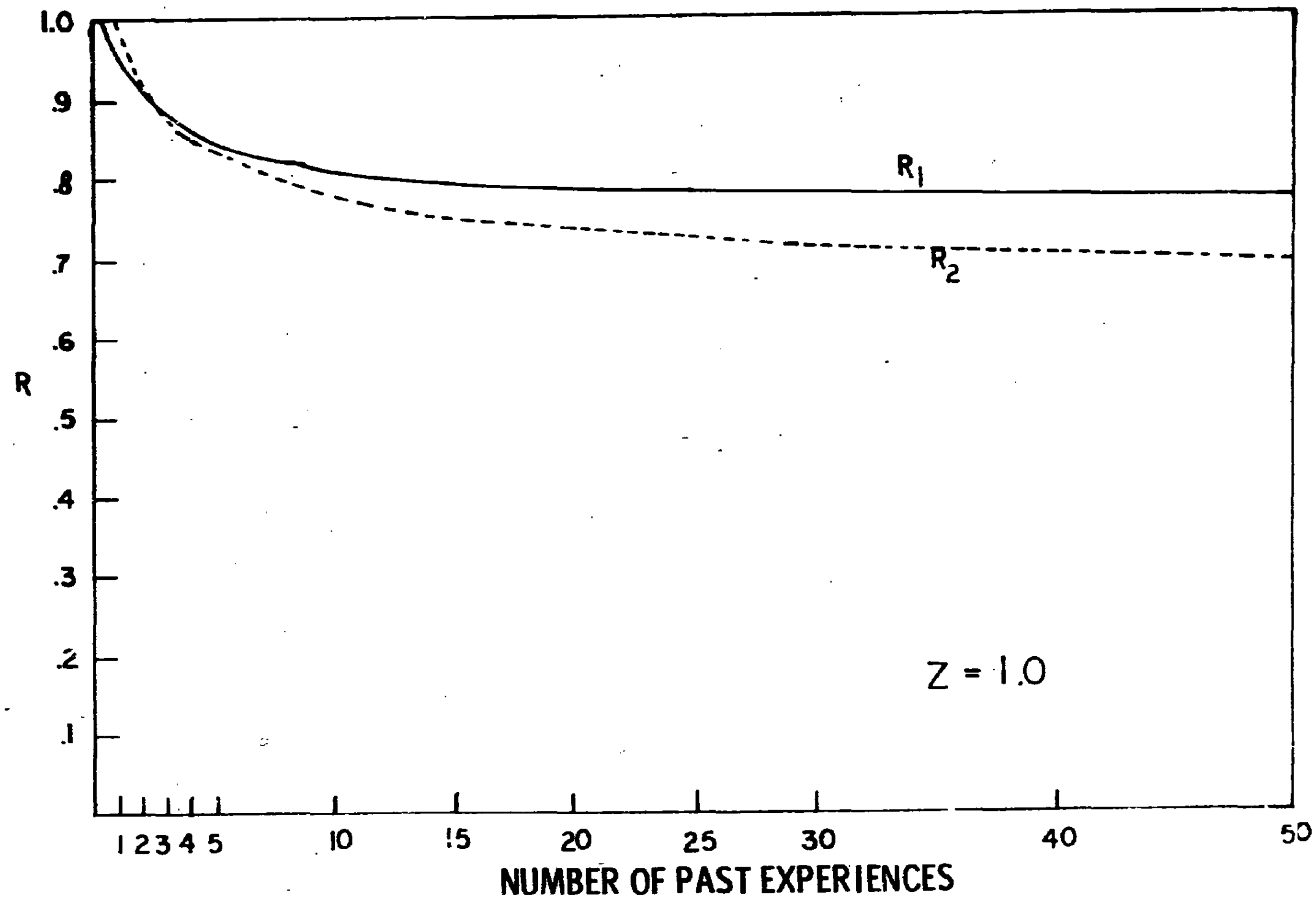


Figure 5
Ratio of cost of Empirical Bayes procedure to the cost of the Classical Procedure in sequential estimation

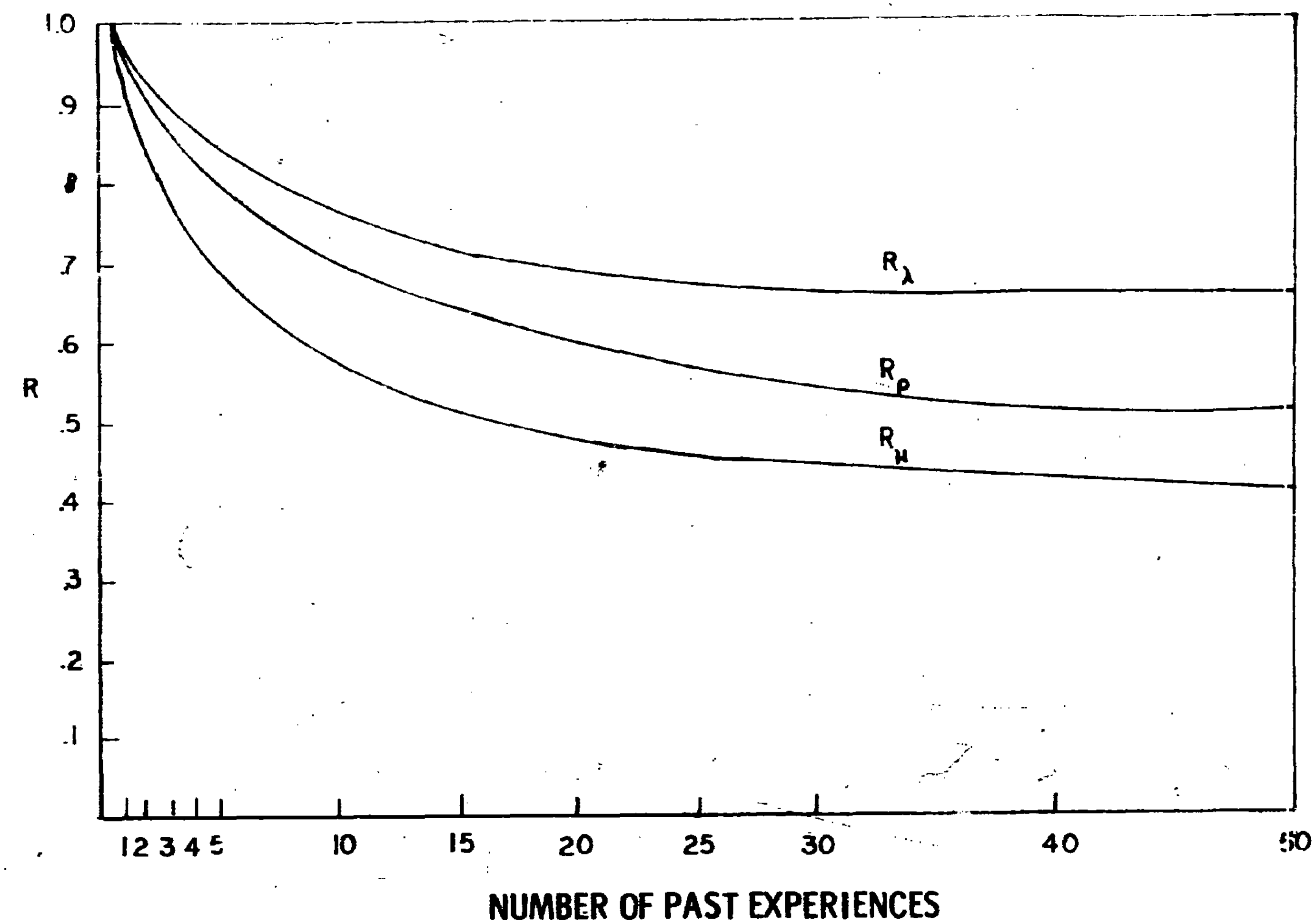
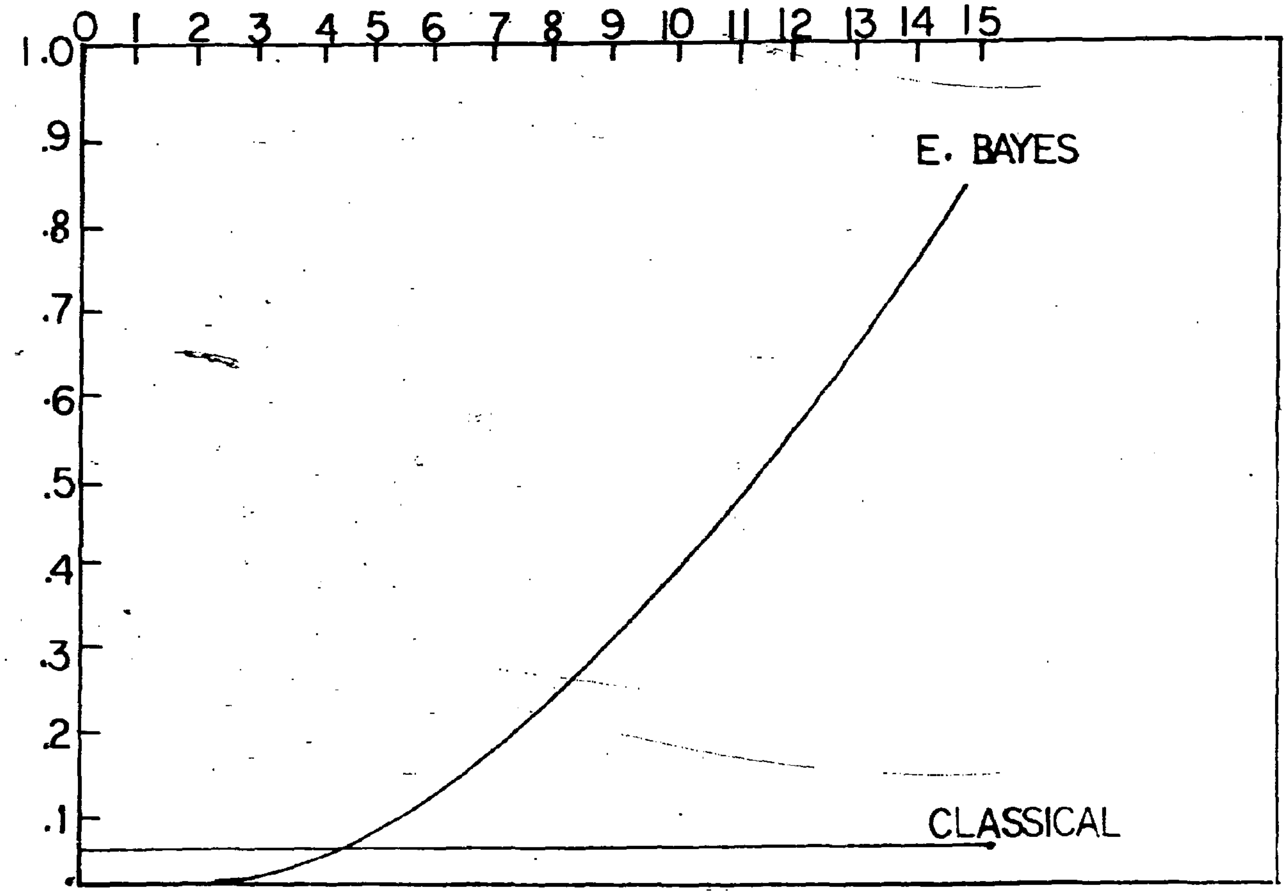


Figure 6
Ratio of mean squared error of the Empirical Bayes Estimation to the MSE of the usual estimation in for an M/M/I Que



NUMBER PAST EXPERIENCES WITH MEAN FRACTION DEFECTIVES P_1

Figure 7

Power curves for 3 control limits when P_0 = .10 and P_1 = .15

FIFTH SAMUEL S. WILKS AWARD

Presentation made by

Dr. Frank E. Grubbs

Preceding page blank

YOUDEN AWARDED THE 1969 SAMUEL S. WILKS MEMORIAL MEDAL

Dr. W. J. Youden, now retired from the National Bureau of Standards, has been awarded the Samuel S. Wilks Memorial Medal for 1969. The announcement of Dr. Youden's selection for the 1969 Wilks Award was one of the highlights of the Fifteenth Annual Conference on the Design of Experiments in Army Research, Development and Testing, which was held at the U.S. Army Missile Command, Huntsville, Alabama, 22-24 October 1969. Dr. Youden has long been recognized as one of the outstanding applied statisticians by both the U. S. A. and countries abroad, as well, having made many fundamental contributions to the design and analysis of statistical experiments and methodology. The citation for Dr. Youden reads as follows:

> To Dr. W. J. Youden, father of 'Youden Squares' and the 'Youden Diagram,' for his extensive contributions to the art and practice of experimentation in the sciences and engineering, through conception and lucid exposition of novel, yet rather elementary, techniques of statistical analysis and crafty application of standard methods; and through his exceptional productivity as an author, indefatigable energy and phenomenal effectiveness as a speaker, by which he has inspired a whole generation of scientists and engineers to greater achievements through application of his unique statistical precepts.

Previous recipients of the Samuel S. Wilks Memorial Medal include: John W. Tukey, of Princeton University (1965); Major General Leslie E. Simon (1966); William G. Cochran of Harvard University (1967); and, Jerzy Neyman of the University of California (1968).

The Samuel S. Wilks Memorial Medal Award is administered by the American Statistical Association, a non-profit, educational and scientific society founded in 1839. The Wilks Award is given each year to a statistician and is based primarily on his contributions to the advancement of scientific or technical knowledge in Army statistics, ingenious application of such knowledge, or successful activity in the fostering of cooperative scientific matters which coincidentally benefit the Army, the Department of Defense, the U.S. Government, and our country generally.

The Award consists of a medal, with a profile of Professor Wilks and the name of the Award on one side, the seal of the American Statistical Association and name of the recipient on the reverse, and a citation and honorarium related to the magnitude of the Award funds. The annual Army Design of Experiments Conferences, at which the Award is given each year, are sponsored by the Army Mathematics Steering Committee on behalf of the Office of the Chief of Research and Development, Department of the Army.

The funds for the S. S. Wilks Memorial Award were donated by Philip G. Rust, Thomasville, Georgia.

With the approval of President A. Ross Eckler of the American Statistical Association, the Wilks Memorial Medal Committee for 1969 consisted of the following:

Professor Robert E. Bechhofer - Cornell University
Professor William G. Cochran - Harvard University
Dr. Francis G. Dressel - Duke University and the Army Research Office-Durham
Dr. Churchill Eisenhart - National Bureau of Standards
Professor Oscar Kempthorne - Iowa State University
Dr. Alexander M. Mood - University of California
Major General Leslie E. Simon - Retired
Dr. John W. Tukey - Princeton University
Dr. Frank E. Grubbs, Chairman - U. S. Army Aberdeen Research and Development Center, Aberdeen Proving Ground, Maryland

BIOGRAPHICAL SKETCH. Dr. Youden was born in Townsville, Australia, on April 12, 1900. Two years later his father returned to his birthplace, Dover, England, with his wife and young son; and the three resided there April 1902 - June 1907. During these years a sister, Dora Alice, and brother, Harry, were born. In 1907, the family of five set out for America, and entered the United States through the Port of New York in July 1907. They lived for a while at Ivoryton, Connecticut, and at Niagara Falls, New York, where Jack attended the local public schools; then they moved to Rochester, New York, in 1916, for Jack's senior year of high school. Youden spent the years 1917-1921 at the University of Rochester, except for one brief interruption to serve his new country as a private in the U. S. Army, October 15 - December 12, 1918. At the University of Rochester, Jack was elected to the National Phi Beta Kappa honor society, and was awarded a B. S. in Chemical Engineering in June 1921. The following academic year, 1921-22, he continued at the University of Rochester as an instructor in Chemistry, then went the two succeeding years, 1922-24, to Columbia University as a graduate fellow in chemistry, earning an M. A. (Chemistry) in 1923; and a Ph.D. (Chemistry), in 1924.

Immediately following receipt of his doctorate, Dr. Youden joined the staff of the Boyce Thompson Institute for Plant Research in Yonkers, New York, as a Physical Chemist. He continued with the Institute in this capacity, with two short leaves of absence and one 3-year assignment as an Operations Analyst with the Army Air Force, until he joined the staff of the National Bureau of Standards in May 1948 as Assistant Chief of the Statistical Engineering Laboratory, which was then beginning its second year of existence.

Dr. Youden was often heard telling a "client" in consultation on statistical aspects of experimentation, or an audience at one of his well attended lectures on statistical methodology, that he is a "chemist," implying, it would appear, that he is really <u>not</u> a statistician. Well, Youden may have been all chemist for his first seven years at the Boyce Thompson Institute, but by September 1931 he had already begun to dish out advice on the statistical aspects of experimentation. The evidence is to be found in his paper entitled, "A Nomogram for use in connection with Gutzeit arsenic determinations on apples," published in Vol. 3, No. 3 of the Contributions from the Boyce Thompson Institute, pp. 363-374. And from impeccable authority we learn that during the academic year 1931-32 he commuted on his own volition from Yonkers to Morningside Heights in New York City to attend Professor Harold Hotelling's lectures on "Statistical Inference" at Columbia University. He was on his way to becoming an expert on statistical aspects of experimentation. From then on he became more and more of a statistician.

The paper that was ultimately to make his name a laboratory, if not a household word, saw publication in early 1937: "Use of Incomplete Block Replications in Estimating Tobacco Mosaic Cirus" (Contributions from Boyce Thompson Institute, Vol. 9, No. 1, pp. 41-48). Here he gave examples and illustrated the application of a new class of symmetrical balanced incomplete block designs that possessed the characteristic "double control" of Latin square designs, without the restriction that the number of replications of each "treatment" (or "variety") must equal the number of "treatments" (or "varieties"). This paper and its new designs led to Dr. Youden obtaining a Rockfeller Fellowship that enabled him to take his first leave of absence from Boyce Thompson, and to devote the academic year 1937-38 to further work in the field of experiment design under the direction of R. A. Fisher himself at the Galton Laboratory, University College, London. Youden's new rectangular experiment designs, termed "Youden Squares" by Fisher and Yates in the introduction to the first edition of their Statistical Tables for Biological Agricultural and Medical Research (1938), were found immediately to be of broad utility in biological and medical research generally; applicable but of less value in agricultural field trials; and with the coming of World War II, Youden Squares proved to be of great value in the scientific and engineering experimentation connected with the research development activities of the war effort of the British and their allies.

Following Pearl Harbor, Dr. Youden took a somewhat longer leave of absence from the Boyce Thompson Institute to serve as an Operations Analyst with the United States Army Air Forces, 1942-45, first as head of the Bombing Accuracy Section of the Operations Analysis Unit of the U. S. Eighth Air Force in Britain, where he directed a group of civilian scientists seeking to determine the controlling factors in bombing accuracy; then, in the latter part of World War II, he was transferred to the Pacific to conduct similar studies preparatory to the B-29 assult on Japan. Stories are legion among the members of the Operations Research Group of the U.S.A.A.F. Eighth Bomber Command about Dr. Youden's exceptional skill in the invention of novel and the adaptation of standard statistical tools of experiment design and analysis to cope with various problems arising in these studies of bombing accuracy. Some of these military applications written up for immediate use, and embalmed for posterity in his ... "How to Improve Formation Bombing," Air Force Manual No. 67, ... 1944, and in the front material to his "Bombing Chart," Air Force Manual No. 79, April 1945. He was awarded the Medal of Freedom in 1946 for his important contributions to the allied victory.

In 1947 Dr. Youden took his third and final leave of absence from the Boyce Thompson Institute: from May to November 1947 he was employed by Project RAND, Douglas Aircraft Company, Santa Monica, California, as a consultant on statistical problems in design and use of military aircraft.

As stated earlier, Dr. Youden joined the staff of the National Bureau of Standards on May 10, 1948, as Assistant Chief of the Statistical Engineering Laboratory, Applied Mathematics Division. Three years later he became a Consultant (on statistical design and analysis of experiments) to the Chief, Applied Mathematics Division, a post that he held until his retirement on June 30, 1965. Since then he has enjoyed the privileges of a Guest Worker at the NBS.

During Dr. Youden's first two years at the NBS, a fraction of his salary was underwritten by the Research and Development Division, Office of the Assistant Chief of Staff, G-4, Department of the Army. This involved coordination with Dr. Merrill M. Flood and others at the headquarters office in the Pentagon; visits to Dr. Ellis Johnson's group at Ft. McNair; and, quite characteristically, Dr. Youden took a number of trips to Army research and development installations in various parts of the country, to size-up "the problems" in their actual habitats.

Dr. Youden's first decade at the National Bureau of Standards saw the invention and publication of his two-sample chart for "Graphical diagnosis of inter-laboratory test results" (Industrial Quality Control, Vol. 15, No. 11, May 1959), now called the "Youden Diagram," which has proved to be an indispensable tool in the inter-laboratory test programs on the National Conference of Standard Laboratories that provide continuing surveillance on the central calibration programs of the U. S. National Measurement System.

The early 1960's saw Dr. Youden's exploitation of a class of selected incomplete block designs of block size two for the specific purpose of identification and estimation of the effects of sources of systematic error, the central theme of his paper, "Systematic Errors in Physical Constants," (Physics Today, September 1961).

Not the least among Dr. Youden's assets are the effectiveness with which he communicates both in writing and speaking; his exceptional productivity in both areas; and the inspiration with which he amuses his readers and audiences. During Dr. Youden's almost two decades at the National Bureau of Standards he was the sole author of thirty, and co-author of fifteen published research papers; the sole author of two books and of seven chapters in other books; and for six years (1954-1959) he authored a highly original bi-monthly column on Statistical Design in the professional journal, Industrial Engineering Chemistry. (These columns have since been brought together and issued in booklet form by the American Chemical Society under the title, "Statistical Design.") During this same period, Dr. Youden gave 211 talks around the country on topics in Statistical Methodology and Experiment Design, under 125 titles, the repetition of some talks being by demand. In addition, he made two lecture tours on behalf of the American Chemical Society, addressed the NBS Scientific Staff Meeting twice, and was called upon repeatedly by the Bureau to address special groups (e.g., high school science teachers). Almost without exception, he was the speaker most highly spoken of afterwards by such audiences.

Dr. Youden's first book, STATISTICAL METHODS FOR CHEMISTS (1951) has had a sale of well over fiteen thousand copies. Together with his "column," this book constitutes one of the best sources of real-life examples of effective applications of statistical principles and techniques in physical-sciences research and development work.

As part of the program of the National Science Teachers Association to place some of the most recent advances in science before junior and senior high school students, Dr. Youden prepared one of the NSTA's Vistas of Science Books, EXPERIMENTATION AND MEASUREMENT (1962). As of July 1969, this booklet has sold over 52,000 copies; and is continuing to sell at the rate of over 1,000 copies per year.

Dr. Youden's total contribution to the art and science of statistics in experimentation is truly impressive. A few weeks before Dr. Youden's retirement from the National Bureau of Standards on June 30, 1965, the Royal Statistical Society elected Dr. Youden to Honorary Fellowship at its annual meeting in London on June 2, 1965. There can be no question that Dr. Youden is a very derserving recipient of the Samuel S. Wilks Memorial Medal for 1969.

THE USE OF A HYBRID COMPUTER TO EVALUATE MAN-MACHINE PERFORMANCE OF COMPLEX VEHICLE CONTROL SYSTEMS

Myrna L. Toivanen, Human Factors Engineering and Simulation, Systems and Research Division, Honeywell Inc., Minneapolis, Minn.

Bernard S. Gurman, Avionics Laboratory, U.S. Army Electronics Command, Fort Monmouth, N. J.

Dr. Erwin Biser, Avionics Laboratory, U.S. Army Electronics Command, Fort Monmouth, N. J.

Summary

This paper describes the role of systematic real-time man-in-the-loop simulations in evaluating man/machine performance of complex vehicle control systems. The simulated system consists of: 1) a hybrid (digital and analog) computer system which is used to simulate the vehicle dynamics and the environmental conditions defining the system state, to drive pilot information displays, and to translate the pilot's control inputs into vehicle responses; 2) electro-mechanical flight displays and/or a computer-addressed cathode-ray-tube (CRT) display which are used to provide the pilot with the information required to perform the defined control task; and, 3) a fixed-base control station which is configured to represent control characteristics of the vehicle being studied. The important considerations in the formulation of the experimental design and schedule are discussed. The performance measures used to evaluate overall system performance are described. The limitations on the application of the simulation study results to the real-world situation are discussed. A summary of the simulation mechanization, methodology, and results of a previous study of helicopter IFR formation flight system requirements is provided as an example of the use of man-in-the-loop simulations to evaluate system performance.

THE USE OF A HYBRID COMPUTER SYSTEM TO EVALUATE PERFORMANCE OF COMPLEX VEHICLE CONTROL SYSTEMS

Introduction

The evaluation of the performance of complex systems would be prohibitively expensive if it were necessary to wait until the system were built and only then test it under actual environmental conditions. The cost involved in redesigning and rebuilding the system at this point in its development would probably increase the cost of production systems by at least an order of magnitude. If the system is man-oriented and there is a chance that system inadequacies may result in catastrophic human injury or loss of life, the risk of system failure is too great and evaluating system performance in this manner becomes unthinkable.

Tests of aircraft or spacecraft systems involve unusually high risks, both in terms of human life and equipment loss. It has been necessary, therefore, for spacecraft and aircraft systems and design engineers to develop alternative techniques for evaluating performance of such systems prior to actual flight tests. The most comprehensive and flexible technique which has been developed has been that of computer simulation.

A computer simulation to be used for system performance analysis consists of an approximate model of both the system and its relevant relationships with its environment mechanized in the form of computer programs (see Figure 1). The system and those environmental conditions which affect system performance are defined in terms of mathematical equations. Then computer programs are written to mechanize models via appropriate computers and algorithms.

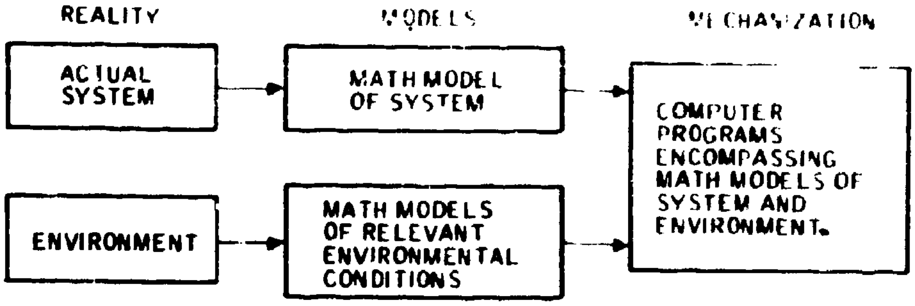


Figure 3. Computer Simulation Development

The simulation of a man-oriented system must include an additional math model for human behavior or must allow the human operator to be included as part of the simulated system. In most complex man-oriented systems the operator's function is a significant factor in total system performance and is too complex to be represented by existing human behavior models. When this is the case, the human operator is included as one of the subsystems of the simulation. Then the simulation must include sensory cues to provide the operator with necessary information and physical apparatus which allows him to perform his function in the simulated system.

In addition to being much less expensive than actual system tests, analysis by simulation is much more flexible. After a simulation is developed, it is possible to systematically investigate system performance resulting from changes in one or a number of the subsystem models or environmental conditions by merely changing the values of the selected parameters in the computer programs. The system design can be changed and tested in much less time and at relatively low cost compared to that involved in rebuilding and reinstalling hardware in the actual system. The system can be tested step by step, analyzing one subsystem while holding other system parameters constant, adding to the complexity of the model only as greater levels of detail are required.

The proper selection of the computers to be used for a simulation depends on the objectives of the study and the nature of the system being simulated. The most appropriate computer is not necessarily that which has the greatest sophistication or capability. Most digital computers are at a point of development now which makes them suitable for use in simulations (i.e., computation speed has been increased to a point where it is suitable for solution of at least the low-frequency system dynamics). The high accuracy, exact repeatability, and flexibility of the digital make it the most appropriate choice for overall simulation control and for simulation of the subsystems which are nonlinear in nature. The analog computer, although not as accurate as the digital, provides continuous solutions and is more appropriate for solving higher frequency system dynamics. It is also more appropriate for simulating characteristics of system hardware which operates in a continuous fashion. Other advantages of the analog for simulation of vehicle dynamics are more operational in nature.. The analog can be programmed to represent linear vehicle dynamics more easily and debugged more quickly than the digital computer. Also, it can be changed on-line without going through the somewhat time-consuming processes of making card changes, recompiling the deck of cards, and reloading the program as required for digital programs. Because of the unique characteristics of the analog and digital computers, a combination of these two computer types, or, a "hybrid" computer system, is extremely attractive for the simulation of real-time, man-oriented systems involving vehicle dynamics.

The extensive use of computer simulation techniques by aircraft and spacecraft systems engineers has made the computer an integral part of the systems design and analysis process. This paper explains how and where the computer, specifically the hybrid computer system, fits into the design and analysis process of man-oriented systems involving vehicle dynamics. Although the discussion will be based on this specific class of systems, most of the techniques and methodology discussed would be applicable to evaluation of a wide variety of complex systems.

Manual IFR Formation Flight System Model

The specific system which will be provided as an example is a manual IFR formation flight system. This system is being investigated by Honeywell under the Joint Army-Navy Aircraft Instrumentation Research (JANAIR) Program, a research and exploratory development program to define and validate advanced concepts which may be applied to future, improved Naval and Army aircraft instrumentation systems. This system would allow a pilot to fly at a fixed position (fixed range and bearing) with respect to another aircraft under poor visibility conditions. The primary elements of the envisioned system (Figure 2a) would be: 1) an on-board special-purpose computer to calculate information regarding aircraft position in the required format for display presentation; 2) a display which will present the necessary information for position control to the pilot; 3) the pilot, who must make control inputs based on the information displayed to him; 4) the controls of the vehicle, which will move the vehicle control surfaces as dictated by the pilot's control movements; 5) the vehicle dynamics, which will result in rotational and translational movements; 6) attitude and/or rate sensors on-board the vehicle which will sense the rotational dynamics of the vehicle; 7) an autopilot, which will provide feedback into the pilot's controls and/or the vehicle dynamics; 8) navigation sensors, which will sense the aircraft's position with respect to the lead aircraft; 9) and a filter which will smooth the data obtained from the sensor. As mentioned previously, the simulation of the system must include relevant environmental conditions; two of these are shown in Figure 2a- i.e., system disturbances, such as turbulent wind conditions (10), and the leader flight profiles (11).

Figure 2b shows a block diagram of the organization of Honeywell's hybrid system in the mechanization of the formation flight system model. Honeywell's hybrid system includes a high-speed digital computer with a real-time capability, a number of analogs, a hybrid link for two-way communication between the digital and analog, a cathode-ray-tube which can be used for generation of

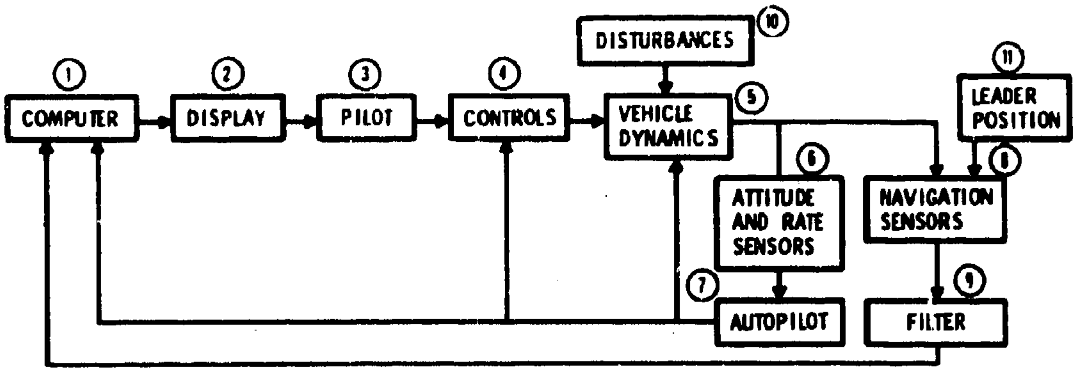


Figure 2a. Manual IFR Formation Flight System Model

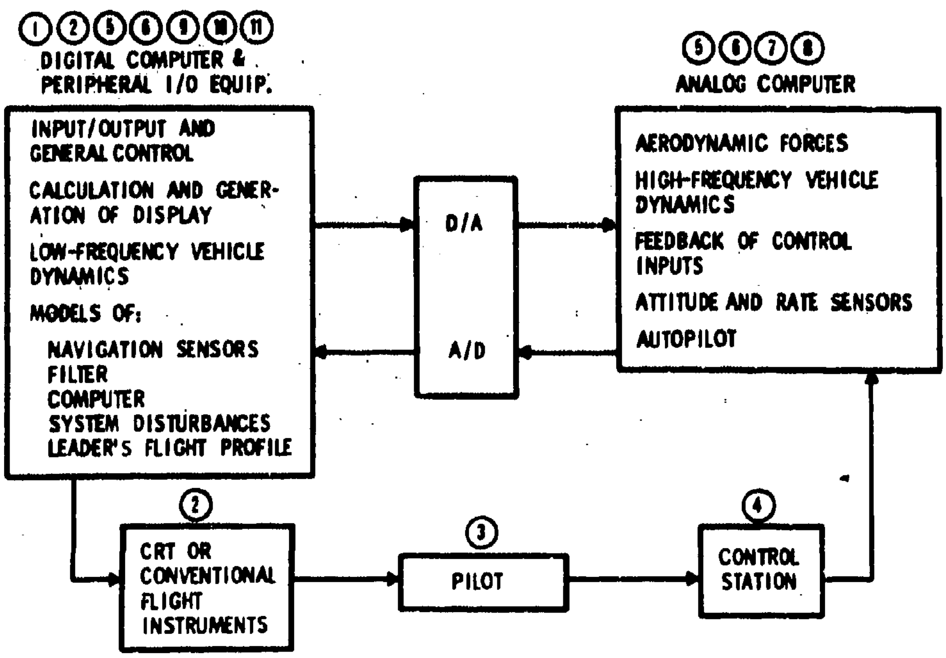


Figure 2b. Simulated IFR FF System

desired display formats (as programmed on the digital), a simulated instrument panel with a number of conventional flight instruments, a fixed-base pilot control station, and standard peripheral equipment which allows communication between the computer and the user. Figure 3 shows the control station located in front of the cathode ray tube display and Figure 4 shows the control station and the simulated aircraft instrument panel.

In the investigations of this simulated formation flight system, pilots have been included in the simulation loop. The flight control and management task which must be performed by the human operator consists of multi-axis aircraft control and three-axis control of position with respect to another aircraft. This task is much too complex to be represented by existing math models of human behavior, which are generally used to represent only single-axis control tasks.

When simulation runs are made, the pilot is seated at the control station in front of a display which provides him with information about his position in space with respect to the lead aircraft. The display can also provide him with aircraft control commands, in which case the position control task becomes primarily one of three-axis tracking. The flight profile of the lead aircraft (programmed on the digital computer) results in the sequential performance of a number of basic maneuvers. The pilot's task is to maintain his position with respect to the lead aircraft throughout these maneuvers. Since this system is to be used for low visibility conditions, the pilot is provided with no visual cues other than the CRT display or specific flight instruments in the simulated cockpit.

System Analysis and Design Process

The process of system analysis and design by simulation and empirical evaluation can be broken down into the following tasks: 1) development of the system

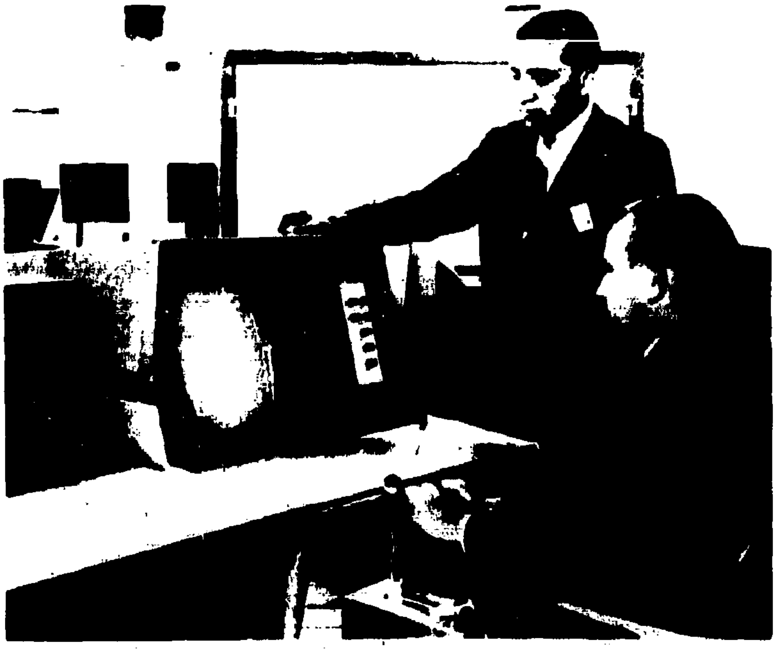


Figure 3. Pilot's Station at Cathode-Ray Tube Display

Figure 4. Pilot's Station in Cockpit Mockup with Electro-Mechanical Flight Instrument Display

model; 2) development of the experimental plan for systematic investigation of system performance; 3) development of the computer simulation; 4) preliminary simulation runs to optimize the simulation; 5) conduct of man-in-the-loop simulations according to the prescribed experimental plan; 6) analysis and interpretation of system performance data; 7) system recommendations and formulation of conclusions. As noted in Figure 5, these tasks are interdependent, with both the sequences and relationships between them being important factors in the effective evaluation of system performance.

One of the mistakes commonly made in the use of simulation techniques is to begin the programming of the computer prior to developing the math models of the system and the experimental plan to be followed in the evaluation of this system. Efficient organization of the program requires at least 90 percent completion of both these tasks prior to development of the simulation programs.

Development of System Model

The definition of the system to be simulated in terms of math models is probably the most difficult task in the system design and analysis process. The difficulty of this task varies with the number of subsystems of the model which have not been previously defined. If the study is a design problem rather than an evaluation of an existing system, the number of such undefined system variables is usually greater. However, even when subsystems are previously defined in terms of hardware characteristics, there are not necessarily existing math models to describe them. Thus the development of a number of math models is usually necessary for any system analysis problem.

The complexity of the simulation to be developed depends on the number of math models required to define the system adequately and the extent of detail necessary for each model. The validity of the simulation, of course, depends on the level of complexity selected for the modeling of the system. Determining the exact level of complexity suitable for a given study is one of the most

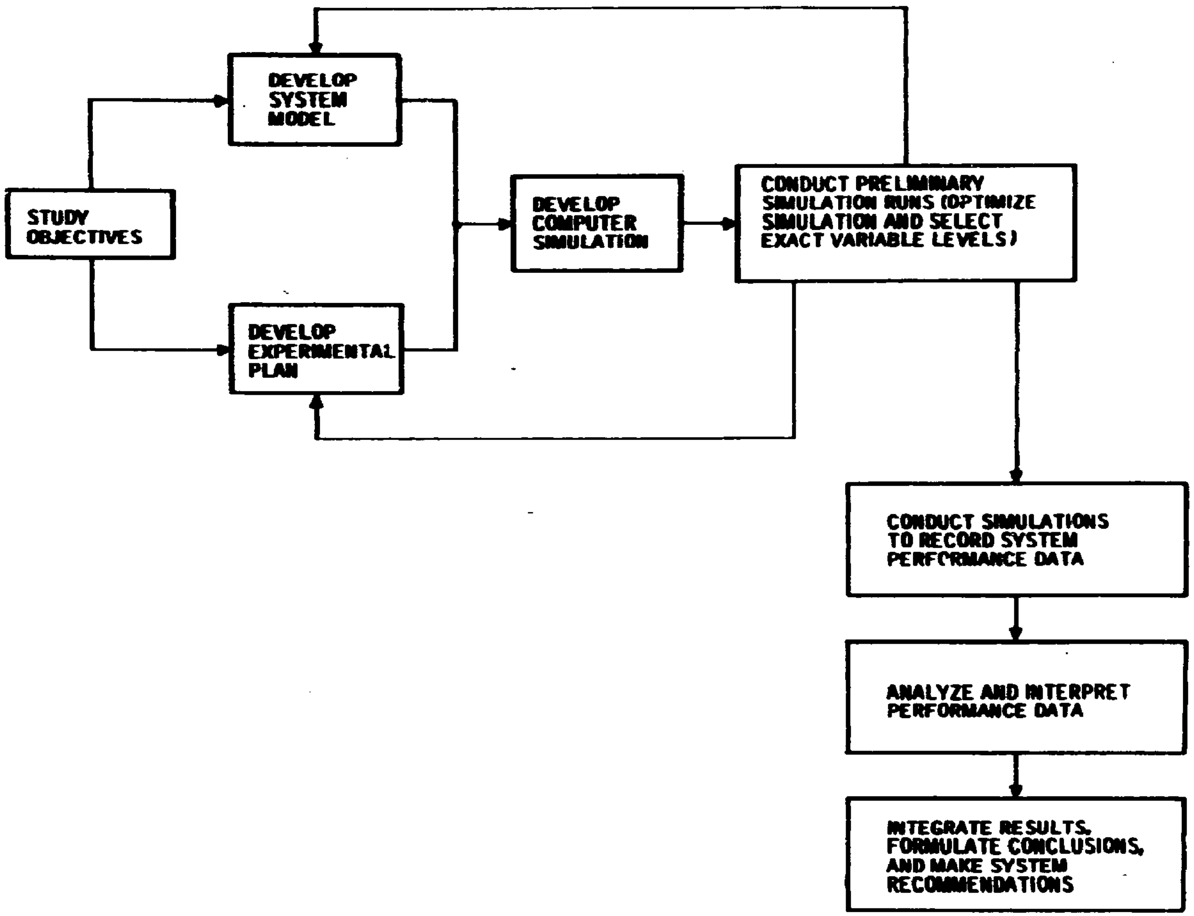


Figure 5. System Analyses and Design Process Using Computer Simulation for Performance Evaluation

difficult chores of the analyst. There must always be a tradeoff between simulation validity and the scope of the study. Determining exactly what this tradeoff should be requires a combination of good judgement and a great deal of previous simulation experience. Unfortunately, as the simulation model approaches exact duplication of the real world, the cost of simulation analysis also approaches the cost of flight testing the actual system.

The subsystems which have been defined for the manual formation flight system model are those shown in Figure 2a (subsystems numbered 1-11). Since the primary objective of this research program has been to develop appropriate cockpit displays for manual IFR formation flight with existing aircraft, the models of the vehicle, the autopilot, and vehicle controls have been predefined at the outset of the investigations. With a human operator as a part of the simulated system, of course, it has not been necessary to develop a math model for the pilot. The remaining subsystem models have been developed during the course of this research program. The greatest amount of time and effort has been devoted to the development of alternative display models. Development of the display model consists of the following tasks:

1) Identifying the information required by the pilot to perform the manual IFR formation flight control task.

2) Incorporating this required information into a total display configuration, which requires:

   a) development of display format (i.e., display symbology) of the primary display,

   b) development of display driving functions (i.e., those equations used to control motion of display elements),

   c) selection of standard cockpit instruments which should be included to provide the required information which could not be incorporated into the primary display.

Two of the display configurations investigated for use in this system are shown in Figures 6 and 7. It may be noted that both display configurations shown provide the pilot with tracking symbols which indicate how he should manipulate the controls to achieve the desired positions. The multi-axis control task required for a high-order control system such as this becomes prohibitively difficult if the pilot must wait for the system's response before receiving feedback regarding the accuracy of his control input. For example, suppose that a pilot makes a roll input to correct a lateral position error. His control input will instantaneously effect a control surface deflection, producing a roll rate, which in turn will result in a significant change in roll attitude within a second or two, which will produce a heading change and lateral rate of movement of the aircraft, which finally, after several seconds, will produce a significant change in the aircraft's lateral position with respect to the leader. In other words, the input response must pass through several integrations before the system finally responds with a change in aircraft position.

To provide immediate knowledge of the results of his control actions, lead information which is based on anticipatory knowledge of the system response must be presented to the pilot. One means of providing this information which has been used successfully in the investigations of this manual IFR formation flight system is display "quickening". Quickening refers to the display of higher-order deriviatives of the system response, which in this case would be time derivatives of the follower aircraft's position errors. Complete quickening, (i.e., presenting the sum of the position error and its derivatives in one element on the display), was utilized to drive the tracking symbols shown in the display formats.

Use of Computer in System Model Development

The computer can be used to aid in developing the system model. Either the digital, the analog, or both can be used for the configuration of a simplified

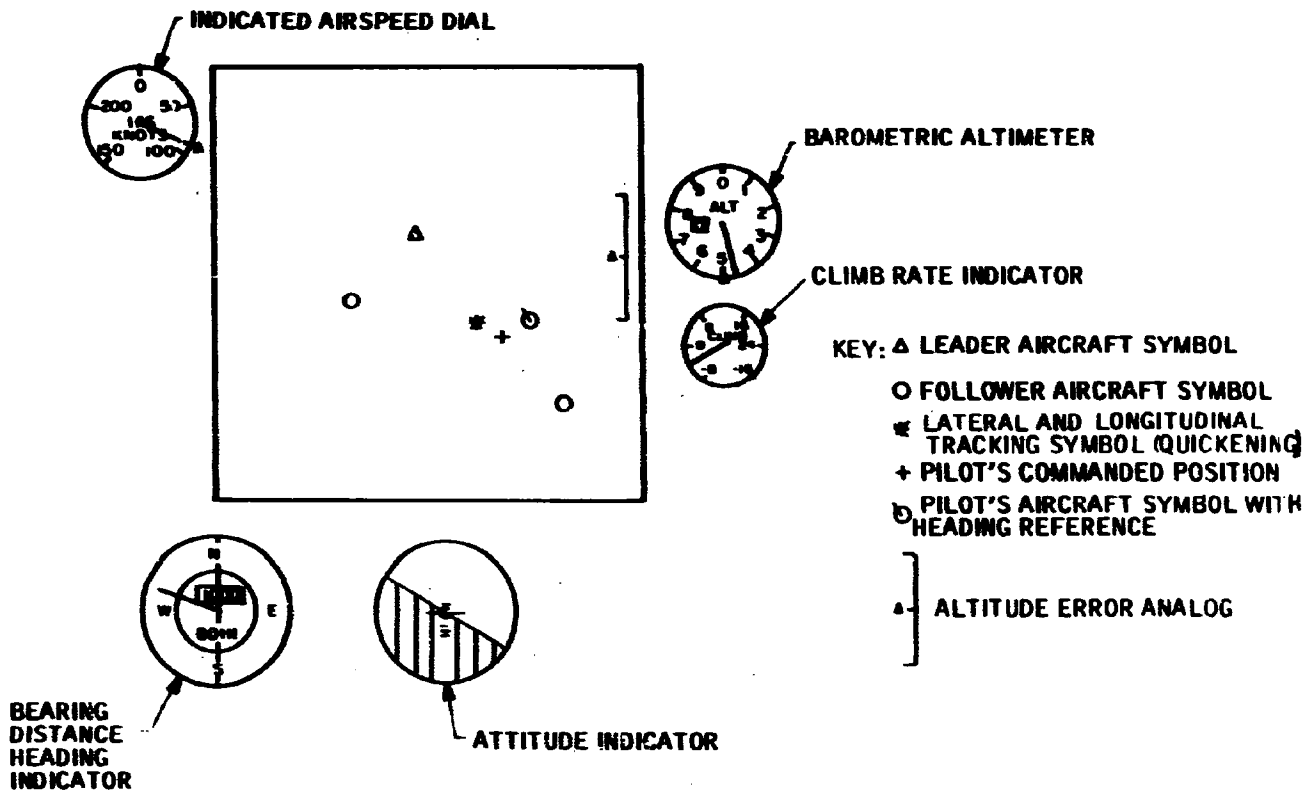
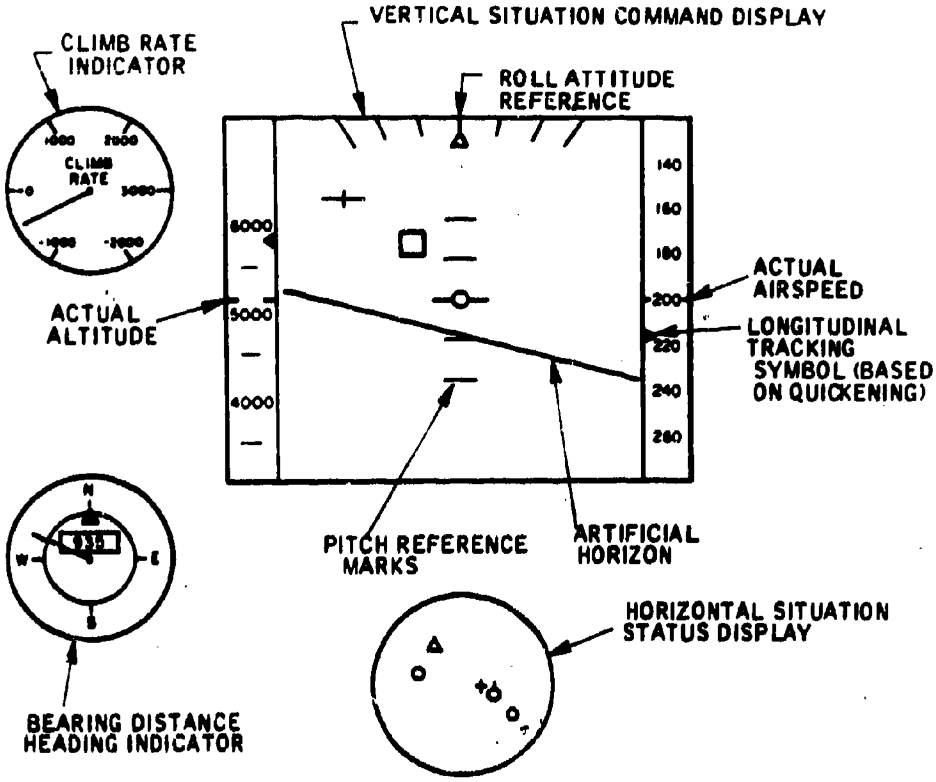


Figure 6. Horizontal Situation Formation Flight Display (Used for the Conventional Helicopter Study)



KEY

VERTICAL SITUATION DISPLAY:

- LATERAL AND VERTICAL TRACKING SYMBOL (BASED ON QUICKENING)
- PILOT'S COMMANDED VERTICAL AND LATERAL POSITION
- PILOT'S OWN AIRCRAFT SYMBOL

HORIZONTAL SITUATION DISPLAY:

- LEADER AIRCRAFT SYMBOL
- PILOT'S COMMANDED LATERAL AND LONGITUDINAL POSITION
- PILOT'S AIRCRAFT SYMBOL WITH HEADING REFERENCE

Figure 7. Vertical Situation Formation Flight Display (Used for the Advanced Rotary-Wing and Jet Fighter Study)

model of the system to empirically evaluate preliminary designs of subsystems. For example, in developing appropriate display driving functions for the display in the manual formation flight system, performance of a simplified system model can be observed to determine which combinations of feedback terms are more appropriate. This simplified model could consist of only one axis of the vehicle control dynamics, a simple model of human behavior (perhaps only a lag) in response to the specified display driving function, and a continuous (analog) recording of vehicle response in terms of position and attitude. Small programs can be written to analytically evaluate specific subsystem math models. For example, to determine the relative velocity required by the follower aircraft to maintain his position with respect to the leader during a turn at various ranges and bearings, the mathematical relationships defining this required velocity are quickly programmed with the ranges and bearings being the variables in the model. Then by merely typing in the desired ranges and bearings, the corresponding relative velocities can be calculated in a fraction of a second.

In addition to the incorporation of the system model, the digital computer programs must include the capabilities of overall simulation control and computer/user communication through appropriate input and output channels. The programs should be organized so that it is easy for the user to change the values of any of the system parameters which are independent variables in the study. They should calculate, record, and output the desired performance measures and be structured to correspond to the experimental procedure to be followed during the course of the man-in-the-loop simulation runs. These additional functions of the computer simulation programs make it necessary that the experimental plan for the study be developed prior to simulation development.

Development of Experimental Plan

The analysis of system performance by simulation, especially when the human operator is included in the simulated system, can involve many hours of simulator runs. If too little time and effort is devoted to planning the experiments, the results of all these hours of simulator runs may turn out to be completely meaningless. No matter how complex or valid the simulation, its usefulness as a tool in evaluating system performance depends on the experimental plan followed in collecting the performance data.

When human performance is one of the contributors to system performance, the system evaluation is based primarily on statistical inference from the performance data collected. The statistical techniques which can be used in interpreting the performance data are necessarily limited by the experimental designs and procedures which have been followed and the exact performance measures which have been recorded. When developing the experimental plan, therefore, it is necessary to select the desired statistical analysis techniques and system performance measures, as well as the experimental design and procedure. The tasks involved and some of the factors which should be considered in developing the experimental plan are shown in Table 1. Specific measures used to describe performance of the manual formation flight system model are shown in Table 2.

Computer Program Development

After the system model and the experimental plan are well defined, it is possible to develop the computer simulation programs in an efficient manner. The digital computer programs are written to incorporate all those subsystem models not programmed on the analog, to calculate and record performance measures, and to allow effective communication between the user and the computer. It should be organized in accordance with the planned

Table 1. Development of Experimental Plan

<u>Identify Independent Variables (IV)</u>

- Will depend on study objectives
- Number of IV's limited by time and money available
- In man-oriented system, human element will be an IV

<u>Identify Dependent Variables (DV)</u>

- Complex system performance evaluation usually requires a number of different performance measures
- When the computer is used to record, calculate, and output performance measures, additional costs associated with additional data are minimal
- Desired statistical analysis techniques should be considered.

<u>Identify Conditions Which are to Remain Constant</u>

- Will include all subsystems of the simulated system which are not independent variables

<u>Determine most Appropriate Experimental Design</u>

- Limited by scope of study
- Should be based on study objectives and nature of the selected IV's

<u>Determine Experimental Procedure to be Followed</u>

- Should minimize learning and order effects
- Should maximize control over experimental constants

<u>Determine Statistical Techniques to be Utilized</u>

- Limited by experimental design and performance measures selected
- Should relate to practical application of study results

Table 2. Performance Measures Used in Analysis of the Manual Formation Flight System

- Mean, standard deviation, and RMS (root-mean-square) error in:
  - Longitudinal position wrt leader
  - Lateral position wrt leader
  - Vertical position wrt leader
  - Range from leader

  These measures are used to indicate the pilot's level of position control precision.

- RMS control stick rates (e.g., $\sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N}}$, where X_i is rate of control stick deflection in inches or degrees per second). This measure is used as an indication of the extent of pilot control activity.

- Collisions with other aircraft

- Continuous time histories of position error, aircraft attitude, and control inputs

- Proportion of time subject utilizes a given display

- Number of attention shifts between displays experimental procedure to allow conduct of the simulation to be as efficient as possible and to free the experimenter from trivial tasks such as timing or data recording. It should be organized such that changing values of system variables can be accomplished easily and quickly.

The analog computer is wired to incorporate those subsystem models which can be more conveniently programmed and debugged on the analog (such as rotational vehicle dynamics) and the control station is configured to represent the controls of the vehicle being simulated.

Once the simulation is developed, it can be used over and over again for a long period of time. It should, therefore, be well-organized and documented so that it can be used and/or modified in the future with minimal difficulty The digital computer programs should be modular in design for ease of checkout and modification. The wiring diagram of the analog simulation should be detailed and complete. The program development should not be considered complete until the program is well documented.

After the entire system model is mechanized in the form of computer programs, the simulation is operational and simulator runs can be conducted.

Preliminary Simulations - System Optimization

Preliminary simulation runs are required to optimize all subsystem models based on man-in-the-loop system performance, to optimize simulation procedures, and to experiment with the ranges of the independent variables of the study. It is usually only after this phase of the systems analysis process that the system model is totally defined. Although the basic math models for the system must be developed prior to writing the computer programs, it is usually desirable to experiment with the parameters which define the exact characteristics of a given model after the simulation is operable. For example, the model of a digital filter may be specifically characterized by parameters which effect differential lags and variance reductions. The optimal values of these parameters are dependent on the specific system application. For complex, man-oriented systems it is usually easier to determine the most appropriate values for such subsystem parameters empirically by means of preliminary simulator runs.

When system response depends on the performance of a human operator, there is often no analytical way to accurately estimate the effect of specific system designs or parameter variation on total system performance. When this is the case, the only way to determine whether a given system model is functional is to test it empirically with a human operator in the simulation loop.

The preliminary simulation phase of the study provides the analyst with rough estimates of system performance. As a result of this preliminary investigation, he can select optimal values for parameters of the subsystem models, select reasonable ranges for the independent variables of the study, and redesign parts of the system if necessary. For example, in the investigations of the manual formation flight system, it was found that the position control task was unreasonably difficult when the display provided only position error information. The analyst experimented with various feedback terms and found that the addition of position error rate and aircraft attitude terms to the position error information made the task much easier. Such preliminary investigations prove to be very valuable in maximizing the effectiveness of the formal experimental tests conducted in the next phase of the system analysis process.

Systematic Man-in-the-Loop Simulation for Collection of Performance Data

After the computer simulation has been completely developed and optimized through preliminary simulator runs, it is finally in the appropriate stage to begin the systematic tests required for performance evaluation. Although conducting real-time man-in-the-loop simulator runs may be the most time-consuming portion of the system evaluation, if preceding phases of the system analysis process have been conducted with care, it is also generally the most straight-forward part of the study from the system analyst's point of view. This phase of the system analysis process consists of familiarizing the subjects with their task under the various experimental conditions and then conducting the simulated runs as prescribed by the experimental plan. For example, in the investigations of the manual formation flight system, the pilot-subjects "fly" the simulated aircraft through the programmed mission repeatedly under each of the experimental conditons until very little further improvement is noted in their position control performance. Then each subject "flies" the required number of missions for the various experimental treatments in the exact sequence prescribed for him by the experimental plan. System performance data are recorded and output for each simulator run.

The most important points to remember in this usually extensive and repetitive process of system testing is that the value of the results depends on strict adherence to the prescribed experimental plan and frequent checks on the accuracy of the simulation programs. The experimenter must guard against becoming lax in his experimental procedure and should frequently, preferably before each simulator run, perform a diagnostic check on the simulation to assure that the programs are working correctly. Although exact repeatability is a characteristic of the digital computer, it is not so for the analog or the link between the digital and the analog. The analog and digital may be improperly connected, some of the potentiometers may be set incorrectly, there may be an amplifier or integrator which is not plugged in securely or not working correctly because of some defect, etc. A diagnostic check can be performed which will usually identify these problem areas quite quickly.

After all prescribed simulator runs have been completed, the system performance data is sorted and collated for interpretation and analysis in the next phase of the system evaluation.

Statistical Analysis and Interpretation of Performance Data

There are numerous statistical techniques which can be employed in analyzing the performance data. The most appropriate techniques depend, of course, on the study objectives and the nature of the system being evaluated. As previously emphasized, the techniques to be used should be selected when the experimental plan and procedure are being defined, to assure that the performance measures recorded and the plan followed will allow valid application of the desired techniques.

The digital computer can be used to perform the lengthy calculations required for the statistical tests. Most large digital computer facilities have a number of general-purpose statistical programs available for performing the more commonly used and well-known statistical tests. A number of such programs are available at Honeywell and have been used in analyzing data obtained in tests of the formation flight system.

When the computer is used both for calculating and recording data and for performing the statistical tests, the added expense associated with recording additional performance measures and/or performing a number of different tests on this data is relatively low. Since it is often difficult for the analyst to determine which performance measures are more appropriate until after the data has been collected and analyzed, it is a good idea for him to record all those measures which seem to be relevant. If digital computer programs are available for performing the desired statistical tests, the same philosophy can be followed in determining the number of tests to be applied to the data. Even when the scope of the study limits the statistical analysis effort, it is important that sufficient performance data be collected and saved. It is often desirable to perform further tests later as additional funding becomes available. Also, it may be that a future study with different objectives requires different methods of data analysis but could be based on the same system performance data. Some of the statistical techniques which have been used in evaluating the manual formation flight system are shown in Table 3.

After results of the statistical tests are obtained, they must be interpreted in terms of their implications in designing or developing the system. Sometimes results which are statistically significant are not significant from a practical systems application standpoint. For example, suppose that the results of a statistical comparison of two displays for the manual formation flight system showed that lateral position control performance was significantly better for one display, and that the average position errors for this display were consistently five feet lower than those for the other display. No matter what the level of statistical significance, this small difference in position error may not be of practical significance in terms of system development. If performance resulting from two such displays were this similar, the system analyst would probably recommend that display which would be less expensive to incorporate into the system. Since the only results of real significance to the system analyst are those which can be related to the design or development of the system, it is important that sufficient time be devoted to interpreting the results of the statistical tests accordingly.

After the analyst has decided what the important study results are, these results should be presented clearly in the final study report. Since the graphical form of presenting data is usually more easily interpretable for the reader than the tabular, it may be helpful to show at least the more important results graphically.

Table 3. Statistical Tests Used to Evaluate Performance of Manual IFR Formation Flight System

Statistical Procedures:	Based on Following Performance Measures:
Factorial analyses of variance	1) RMS position error 2) RMS control stick rate 3) Mean position error
Calculation of mean and/or median position errors for factorial combinations of independent experimental variables.	Mean position error for a specific subject, treatment, and maneuver.
Calculation of mean and/or median variances for factorial combinations of independent experimental variables.	Standard deviations around the mean position error for a specific subject, treatment, and maneuver.
Calculation of correlation and/or regressions between system variables.	E. g., between lateral and longitudinal position errors, or between fore-aft and right-left control stick movements, etc.

Formulation of Conclusions and System Recommendations

The final step in the system evaluation process is the careful examination of study results to formulate conclusions and make system recommendations. Sometimes in the early phases of the system investigation, recommendations regarding system implementation--such as required hardware characteristics--cannot be made until further research has been conducted. The study conclusions at this point usually relate primarily to the feasibility of system concepts. The primary system recommendations made will be suggested areas for further investigation, utilizing those effective system concepts and models which have been developed during the previous system studies as a basis for future studies. The system analysis process described in this paper would then be repeated a number of times until sufficient system aspects have been investigated to allow recommendations regarding hardware characteristics and specifications.

Results and conclusions of the formation flight system research are provided below as an example of the steps which must be completed prior to actual system development and the type of conclusions which can be drawn from system evaluation studies based on computer simulation analysis.

Results and Conclusions of Research on Manual IFR Formation Flight System

The research on the formation flight system has not yet reached the point where the exact specification of required hardware characteristics would be desirable. The philosophy followed in this system investigation has been to first determine the feasibility of basic system concepts and associated system performance, assuming no limitations imposed by system hardware. Then, one by one, the system limitations imposed by realistic hardware characteristics and expected environmental conditions have been added and resultant system performance evaluated. Since there are so many system variables in a complex system such as this, it has not been desirable, either from a cost or operational standpoint, to investigate all these variables simultaneously. Instead, a number of different studies have been conducted and each new study has utilized results of previous studies as a basis for evaluating additional system variables. The studies which have been performed to investigate the manual IFR formation flight system and the major results of these studies are described briefly below.

The first study (Reference 1) of this program was conducted to investigate basic information requirements for the manual IFR formation flight control task for rotary-wing aircraft and to evaluate display concepts for a computer-addressed cathode-ray tube display. The effects of turbulent wind conditions and subsidiary pilot workload were also investigated in this study. This study assumed no limitations on sensor outputs (i.e., high data transmission rate and no measurement noise). Some of the study results are shown in Figures 8 through 11. The major conclusion of the study was that formation flight under IFR conditions appears to be a realizable goal with the aid of the computer-generated display formats developed.

The second study (Reference 2) was conducted to evaluate an existing helicopter formation flight system. This study assumed the sensor, computer, and display characteristics of the system being evaluated. Results of this study demonstrated the important effects of filtering techniques, data transmission rate, and display driving functions on total system performance (see Figures 12 through 14).

A third study (Reference 3) was conducted to evaluate the effectiveness of conventional flight instruments in a manual IFR helicopter formation flight system. Two state-of-the-art electro-mechanical displays, i.e., a flight director and a horizontal situation indicator, were used in conjunction to display the required information and were evaluated under alternative display formats. This study again assumed no limitations on sensor system outputs.

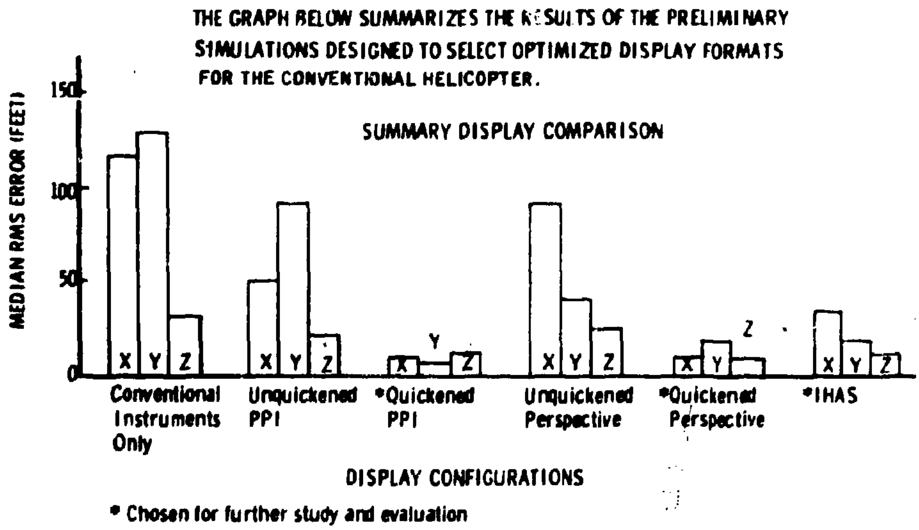


Figure 8. Original Study of Concept Feasibiltiy and Display Requirements for the Conventional Helicopter

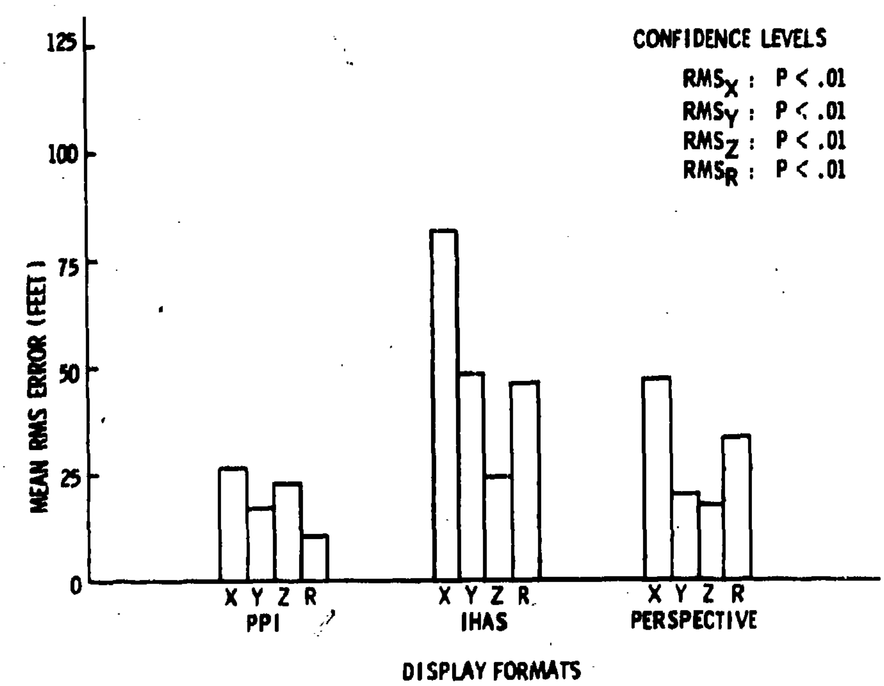


Figure 9. Comparative Evaluation of Display Formats for Conventional Helicopter Study

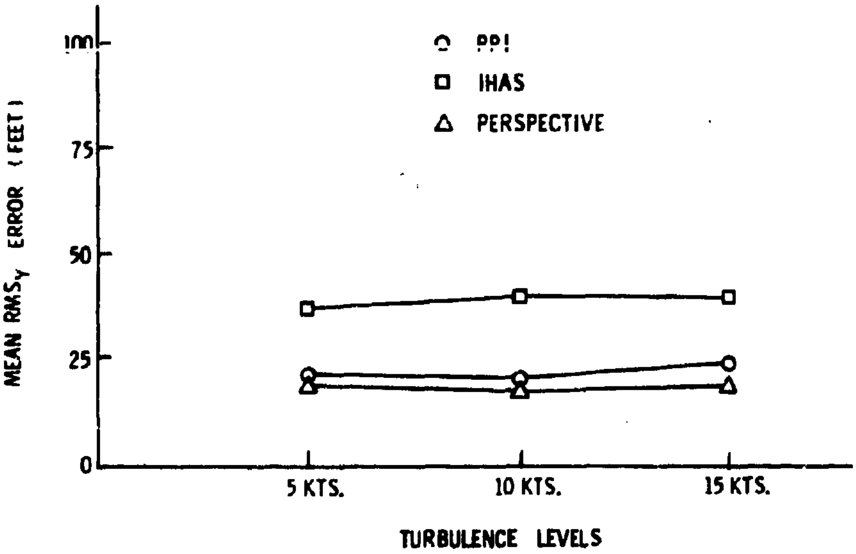


Figure 10. Effects of Turbulence on System Performance - Conventional Helicopter Study

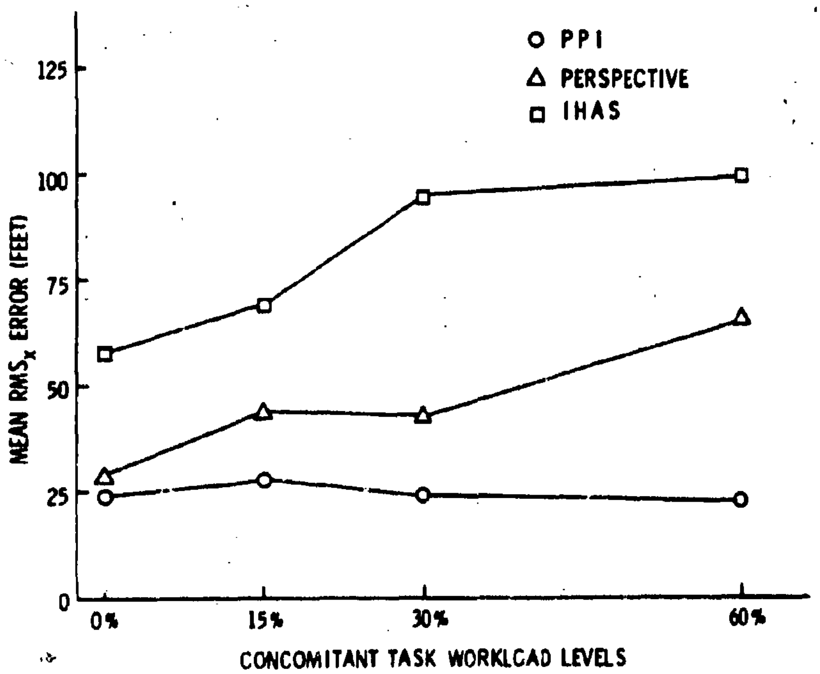


Figure 11. Effects of Subsidiary Pilot Workload on System Performance - Conventional Helicopter

THE GRAPH BELOW SHOWS MEAN RMS ERRORS PLOTTED AS FUNCTIONS OF THE SYSTEM UPDATE TIME INTERVAL. SYSTEM UPDATE RATE IS THE RECIPROCAL OF SYSTEM UPDATE TIME INTERVAL.

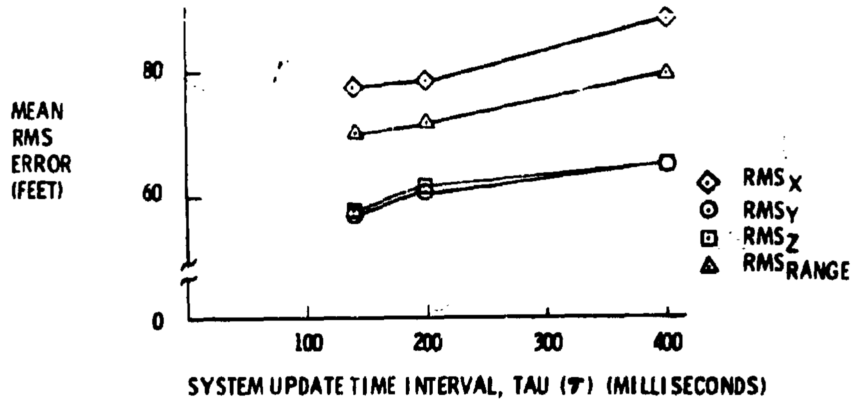


Figure 12. Sample Prototype System Evaluation - Effect of System Update Rate on System Performance - Conventional Helicopter

THIS GRAPH SHOWS MEDIAN RMS ERRORS AS A FUNCTION OF THE GAINS USED ON THE POSITION ERROR RATES IN THE DISPLAY HORIZONTAL QUICKENING EQUATIONS.

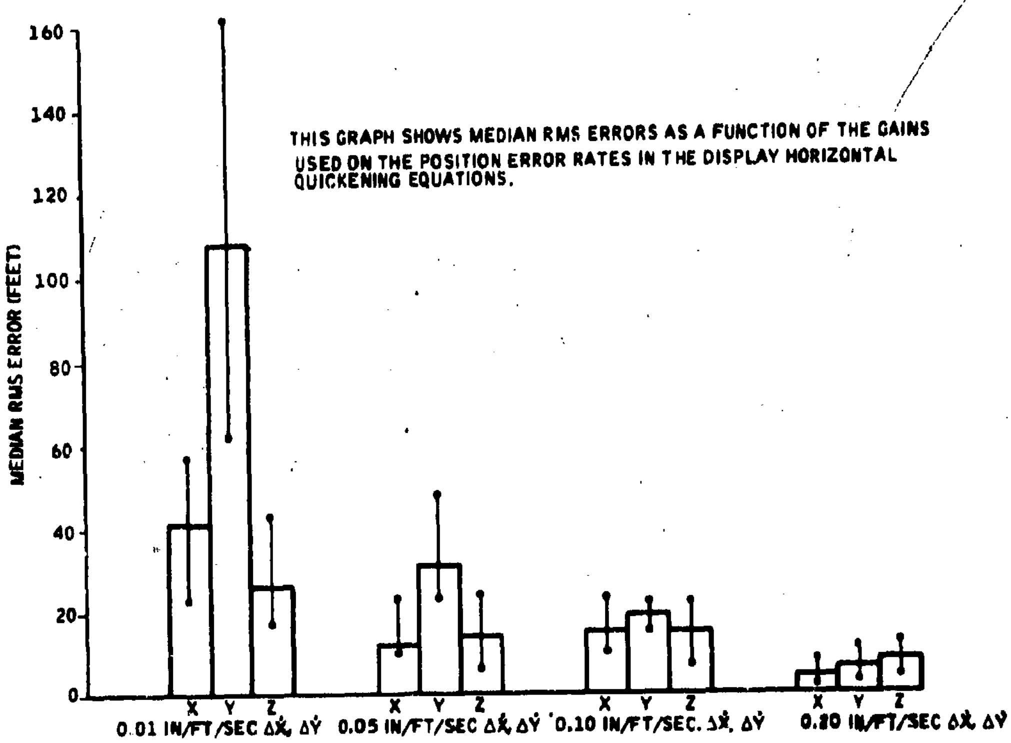


Figure 13. Sample Prototype System Evaluation - Effects of Terms in Display Quickening Equations on System Performance

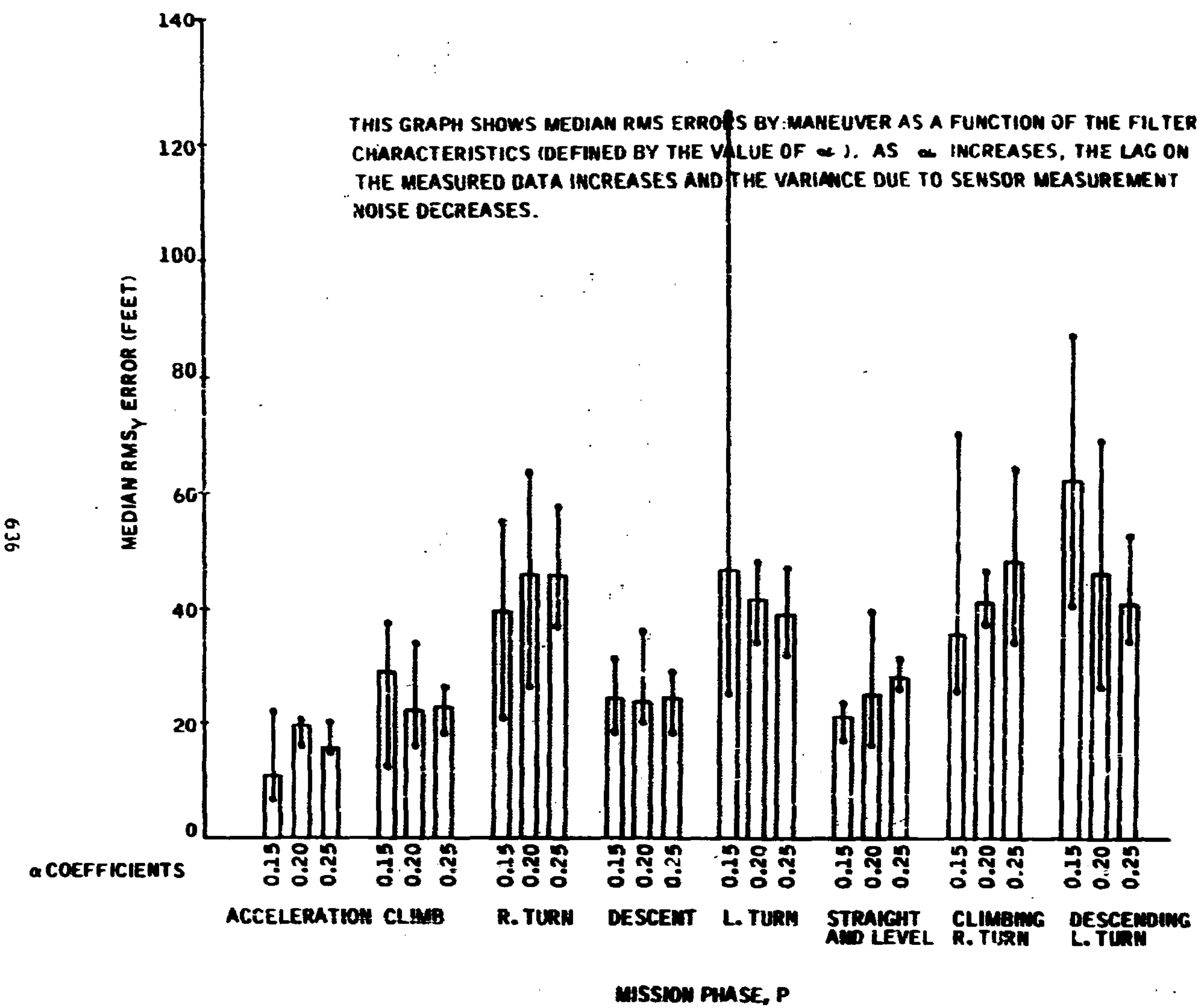


Figure 14. Sample Prototype System Evaluation - Effect of Filter Characteristics - Helicopter Study

The results of this study (see Figures 15 and 16) indicated that it was possible to obtain position control with the electro-mechanical displays comparable to that obtained with the computer-generated displays.

The fourth study (Reference 4) investigated the relationship of variations in the effective data transmission rate (i.e., defined as the update rate of the information presented on the displays) and the effective level of measurement noise (i.e., noise which appears on the display after filtering) on pilot/system performance in the helicopter IFR formation flight mode. Results (see Figures 17 through 19) showed: 1) that position control performance degraded with increasing measurement noise; 2) that position control performance did not improve significantly by increasing the update rate above 4/second; and 3) that optimal display driving functions and data filtering techniques are dependent on the data update rate and accuracy characteristics of the system.

The objective of the fifth study (Reference 5) was to evaluate the effect of varying levels of automatic control assistance on pilot/system performance in the simulated helicopter IFR formation flight mode. An information update rate of 4/second and a moderate noise level were assumed throughout this study. Results of the study (see Figures 20 and 21) indicated that increasing the level of automatic control assistance provided greater system stability and made the pilot's control task less demanding, but did not significantly improve position control performance over that obtainable manually with the aid of the quickened display.

The sixth study (Reference 6) was conducted to define information and display requirements and investigate variable sensor output characteristics for two additional vehicle classes, i.e., the advanced rotary-wing and the jet fighter aircraft. The results (see Figures 22 through 25) indicated that manual IFR formation flight with the envisioned system appears to be feasible for the advanced rotary-wing and the jet fighter aircraft and that the effects

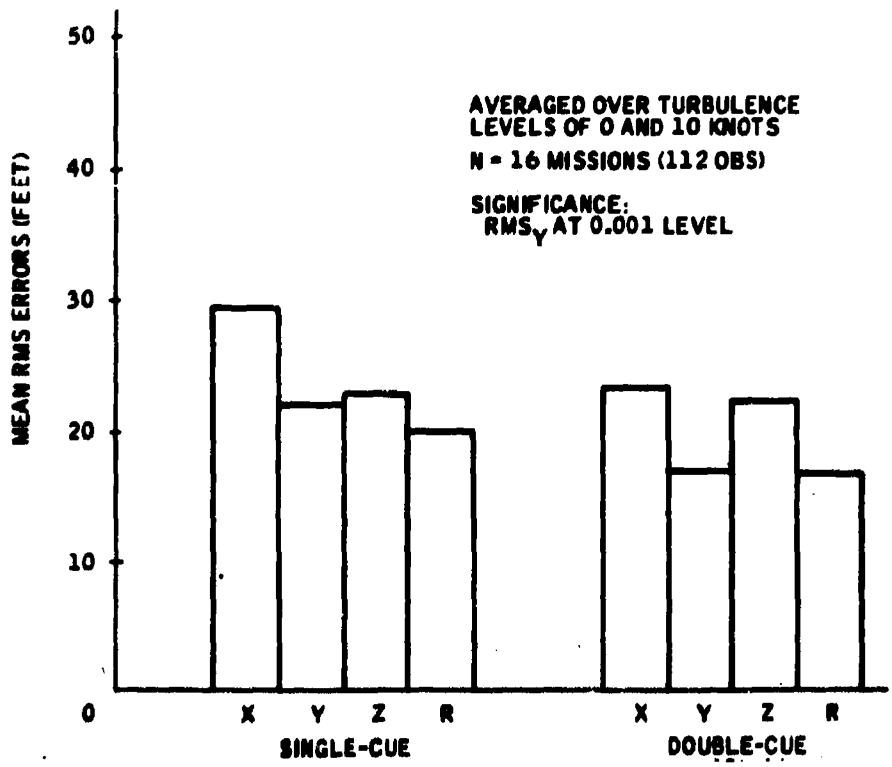


Figure 15. Comparative Evaluation of Two Flight Director Displays for Conventional Helicopter

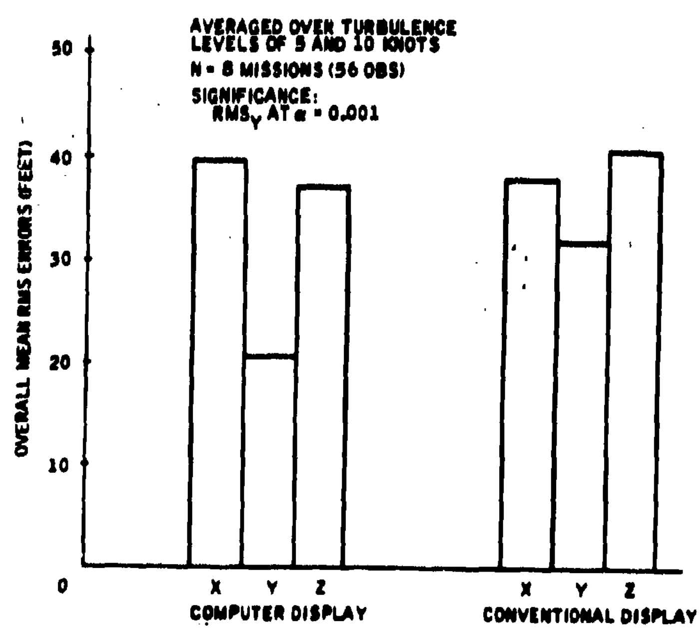


Figure 16. Computer-Generated Display versus Conventional Flight Director Display - Conventional Helicopter

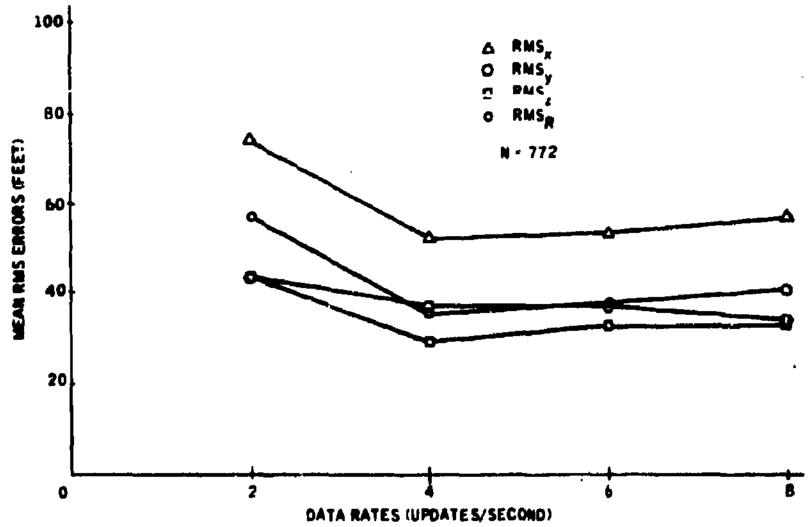


Figure 17. RMS Position Errors by Data Rate - Conventional Helicopter

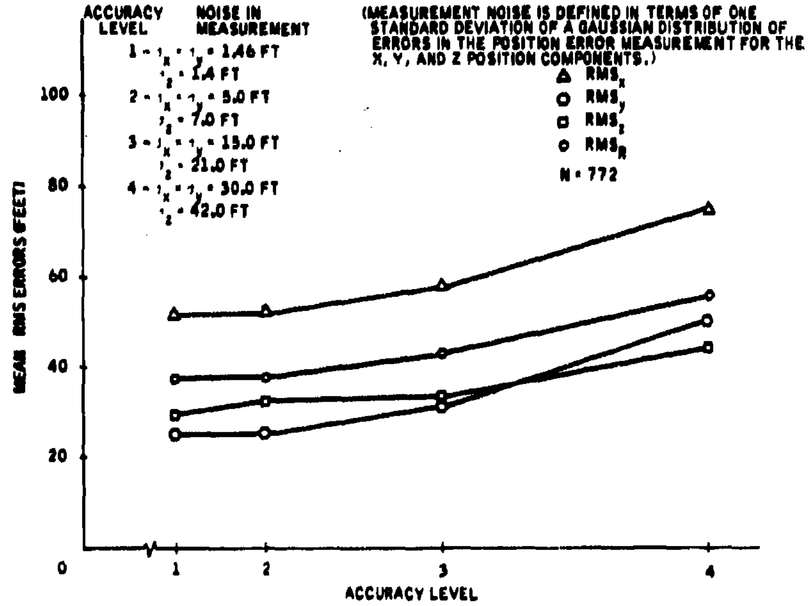


Figure 18. RMS Position Errors by Data Measurement Noise Level - Conventional Helicopter

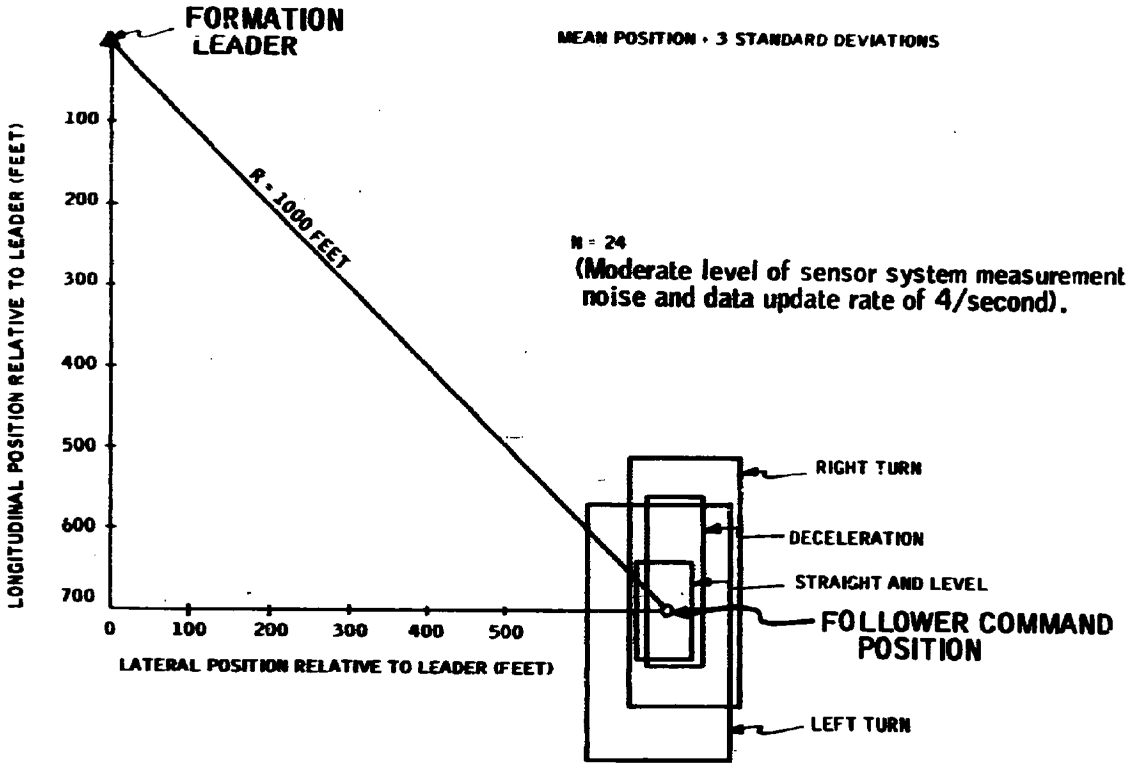


Figure 19. Horizontal Position Error Envelope for Conventional Helicopter

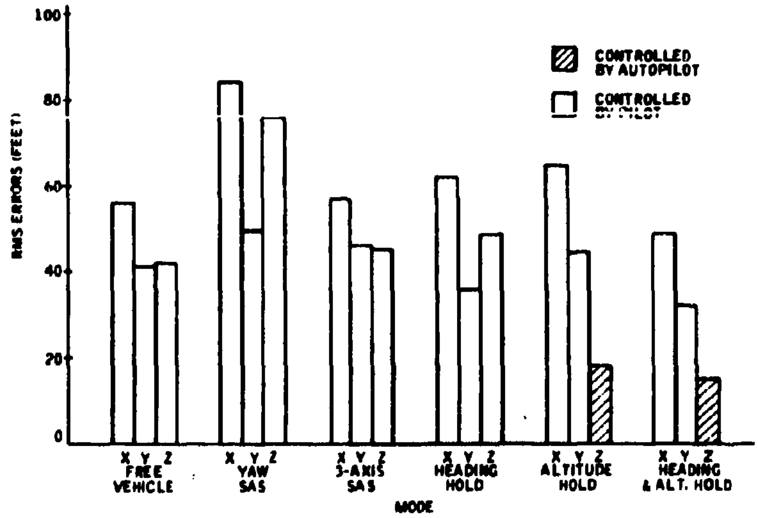


Figure 20. Mean RMS Errors by Autopilot Mode for Conventional Helicopter

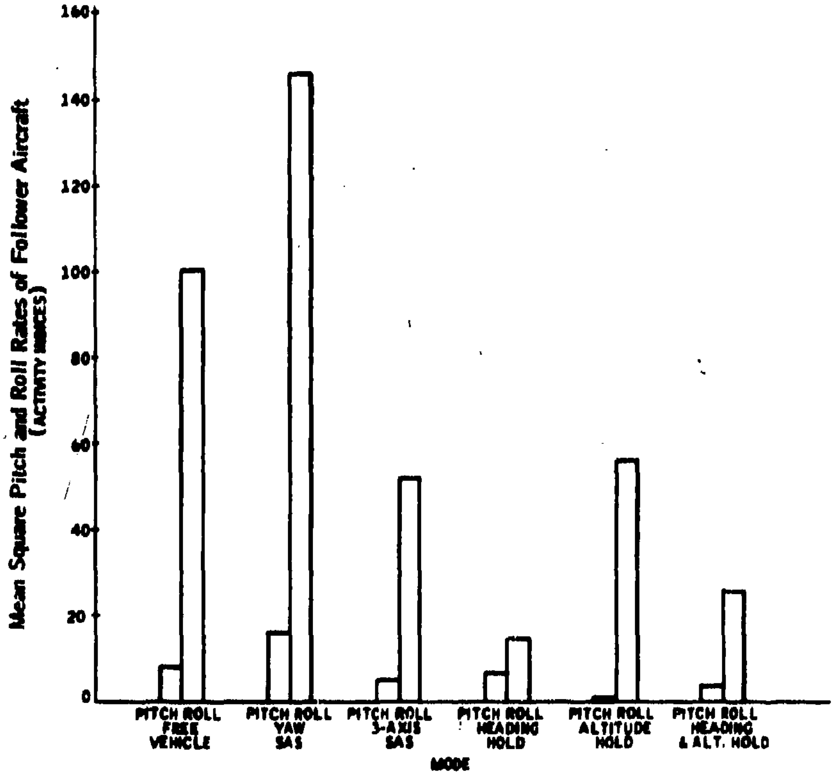


Figure 21. Aircraft Pitch and Roll Activity as a Function of Level of Automatic Assistance - Conventional Helicopter

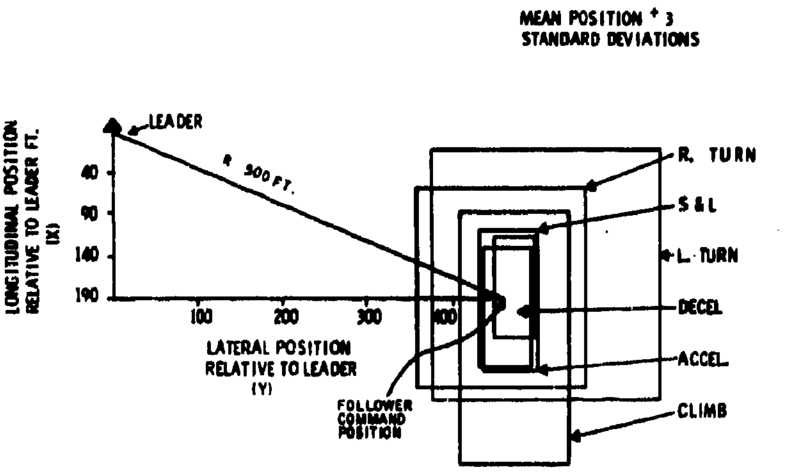


Figure 22. Horizontal Error Envelope by Maneuver for the Advanced Rotary-Wing

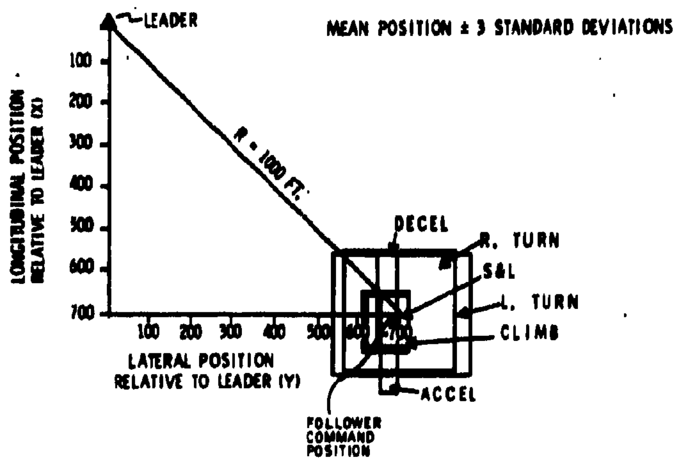


Figure 23. Horizontal Position Error Envelope by Maneuver for the Jet Fighter

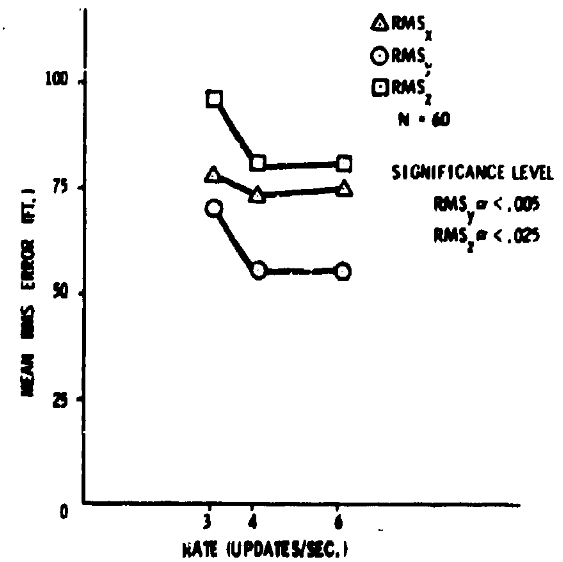


Figure 24. Effects of Data Update Rate on System Performance - Jet Fighter

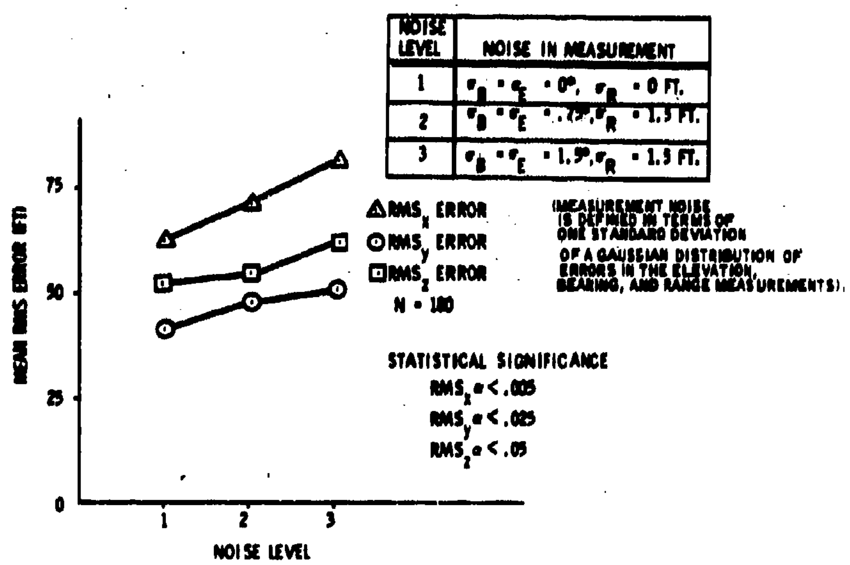


Figure 25. Effects of Sensor Measurement Noise on System Performance - Advanced Rotary-Wing

on system performance of variation in measurement noise and display update rate were the same as found previously for the helicopter.

Some of the basic conclusions which have been drawn as a result of these research studies are summarized below.

Display driving functions are more important than the display format in determining the pilot's control performance. The results of the various display evaluations suggest that as long as the basic display/control relationships are satisfied, all required information is presented, and the display formats are interpretable, a number of different display formats (in terms of specific symbology and orientation) are appropriate for the envisioned IFR formation flight system.

Manual IFR formation flight with the system as modeled for these research studies appears to be a realizable goal for the conventional helicopter, the advanced rotary-wing, and the jet fighter. The level of position-control precision to be expected for a specific aircraft class, given the presentation of all required information, the use of an appropriate display format, and optimal display quickening, will be a function of at least the following variables:

1) Rate at which new information is available for display
2) Level of measurement noise on positon information.
3) Filtering technique used to smooth the noisy data
4) Experience level and capability of the individual pilot
5) Extent of pilot workload required for subsidiary tasks.

It is suspected that other variables, such as pilot fatigue, aircraft separation distances, formation geometry, formation airspeed, command mode (i.e., fly-to or fly-from), lead aircraft perturbation, etc., also effect postion-control performance. Current research under the JANAIR program includes investigation of a number of these additional system variables. Prior to exact specification of the hardware required for the manual IFR formation flight system, these and other seemingly relevant system variables should be investigated.

The Value of Simulation in System Performance Evaluation

The conclusions summarized above are somewhat tentative in nature and do not specify the exact level of system performance which can be expected with the developed system under actual flight conditions. It must be emphasized that simulation analysis can not be a substitute for actual system tests. Rather, simulation is a compromise method of evaluating system performance with a degree of validity which falls somewhere between preliminary pencil-and-paper analyses and tests of an actual breadboard system. Exactly where it falls in terms of validity depends on the complexity of the system model which is developed for the simulation.

The value of simulation as a system evaluation tool is greatest for those systems which cannot be easily defined analytically (such as those involving complex human behavior) and which cannot be tested under real-world conditions without extremely high risks (i.e., in terms of human life and/or system costs). For this type of system, compromise methods of system evaluation are necessary, and they are not meant to replace actual system tests, but to precede and minimize the extent of these tests.

Although there are inherent constraints and limitations in the evaluation of system performance by simulation analysis, it is extremely useful in answering questions such as the following:

- Is a given system design or concept feasible?
- How does one method or design compare to another?
- What are reasonable minimum and maximum limits on the variation of a given system parameter?
- What are the effects of varying two or more system parameters simultaneously?
- Can the human operator perform the required task?
- Which system tasks should be performed automatically?
- What is the general relationship between system performance and a given system parameter or environmental condition?

If the system analyst conducts his system investigation scientifically and if he tempers his formulation of conclusions and system recommendations by acknowledgement of the constraints and limitations of the simulation, he will find computer simulation techniques to be an invaluable tool in the system analysis and design process.

GLOSSARY

Activity indices	Mean-square pitch and/or roll rates of the simulated aircraft
A/C	Aircraft
Autopilot mode	Level of automatic control included in the aircraft control system. Ranges from simple damping in a single axis of the aircraft to control of the aircraft's heading and altitude.
Concomitant task	A class of secondary task which is performed simultaneously with the primary task and is highly quantifiable in nature. In the study referenced in this paper, the primary aircraft position control task was performed continuously and the pilot's formation flight display was intermittently blanked out. Then secondary task cues were provided, requiring that the pilot simultaneously perform the secondary task at a forced-pace (i.e., the frequency and time interval of the formation flight display interruption was controlled by the experimenter rather than left to pilot discretion). The levels of the secondary workload were defined in terms of percentage of time the pilot was required to perform the secondary task.

Data transmission and/or update rate	The rate at which new information about the follower aircraft's position in space (with respect to the lead aircraft of the formation) is available for display.
Dependent variables	Variables of an experiment which describe system performance. These performance measures are assumed to reflect changes in the levels of the independent variables of the experiment and are thus considered to be "dependent".
Double-cue flight director	Electro-mechanical flight instrument currently used in both fixed-wing and helicopter aircraft to provide information about aircraft pitch and roll attitudes. The "double-cue" flight director has two separate moving elements, one representing pitch and the other roll attitude deviations.
Filter	In the referenced studies the filter simulated was a digital α-β filter. The filter is required to smooth the aircraft position information obtained from the assumed sensor system. Sensor systems are usually characterized by a certain level of measurement noise.
IFR	Instrument Flight Rules -- This term is used in this report to refer to very low visibility conditions when a pilot would have to depend primarily on instruments for visual cues.

IHAS display format — Integrated Helicopter Avionics System - The vertical situation display format which was intended as a part of this avionics system was configured for use in the IFR formation flight mode in one of the referenced studies.

Independent variables — Parameters of a system which are varied to investigate their effect on system performance.

Mean position error — $\sum_{i=1}^{N} X_i/N$, where X_i was the measured position in a specific axis during a specific maneuver and N was the total number of position measurements recorded during the maneuver. For the referenced studies N was equal to the total maneuver length times the display update rate.

Measurement noise — Used to refer to the error in the measurement of the follower aircraft's position with respect to the leader. The sensor system was assumed to include noise with a normal (Gaussian) distribution described by the standard deviation. In the referenced studies these standard deviations were defined in terms of the bearing (σ_B), elevation (σ_E), and range (σ_R) measurements (in degrees), or in terms of the longitudinal (σ_X), lateral (σ_Y), and vertical (σ_Z) position (in feet).

Perspective display format	A display format configured for the formation flight mode which provided the pilot with a three-dimensional representation (in two dimensions) of his formation position with respect to the leader.
PPI display	Plan Position Indicator - A display format configured for the formation flight mode which presented a horizontal view of all aircraft in the formation.
Quickening	A method of providing lead or anticipatory information regarding the system's response. As used in the referenced studies, it consisted of adding higher order derivatives (i.e., rate of change of position) of the system's response (change in aircraft position) to the actual position error. The resultant sum was used to drive one element on the display.
RMS	Root-Mean-Square-- $\sqrt{\sum_{i=1}^{N} X_i^2/N}$, where X_i was the measured position in a specific axis during a specific maneuver and N was the total number of position measurements recorded during the maneuver. The RMS measure was also used to represent the levels of pilot and aircraft control activity during a given maneuver, in which case X_i represented respectively the rate of movement of the control stick or the aircraft's attitude.

SAS	Stability Augmentation System.
Single-cue flight director	An electro-mechanical flight display which has only one moving element to represent both pitch and roll attitudes of the aircraft. The element represents both aircraft axes by movement in two different axes on the display.
SK/FF	Stationkeeping/Formation Flight.
Standard deviation	$\sqrt{\sum_{i=1}^{N} (\bar{X} - X_i)^2 / N}$, where $\bar{X}$ is the mean value of the observations (X_i) and N is the total number of observations.
Subsidiary pilot workload	The pilot's workload on tasks other than his primary task (aircraft position control in the referenced studies). See the description of "concomitant task" for more detail on how this subsidiary workload was simulated.
System update time interval	Reciprocal of the rate at which the pilot received new information about his position with respect to the formation leader for display.
α	Coefficient of α-β digital filtering model which determines weight of current raw position measurement versus average over old measurements.
β	Coefficient of α-β digital filtering model which determines weight of current velocity measurement versus average over old measurements.

REFERENCES

1) "Display Requirements Study for Helicopter IFR Formation Flight", Honeywell Document No. 12543-FR1, AD 665-277, JANAIR Report NR 213-054, January 1968.

2) "Experimental Evaluation of a Formation Flight System", Honeywell Document No. 12543-IR5, JANAIR Report 6804002, July 1968.

3) "Evaluation of Conventional Flight Instruments for Use in Helicopter IFR Formation Flight", Honeywell Document No. 12543-FR2, JANAIR Report 680204, February 1968.

4) "Effect of Variations in System Data Rates and Measurement Accuracies on Pilot Performance in the Helicopter IFR Formation Flight Mode", Honeywell Document No. 12543-FR3, JANAIR Report 680408, April, 1969.

5) "Effects of Varying Levels of Autopilot Assistance on Pilot Performance in the IFR Formation Flight Mode", Honeywell Document No. 12543-FR4, Preliminary JANAIR Report, June, 1968.

6) "IFR Formation Flight Display/System Requirements for Advanced Rotary-Wing and Jet Fighter Aircraft", Honeywell Document No. 12543-FR5, Preliminary JANAIR Report, June, 1969.

EXPERIMENTAL DESIGN CONSIDERATIONS IN VALIDATING A METHOD OF MODELING A MAN-ORGANIZED SYSTEM

Bob B. Lukens and Robert A. Brown
University of Alabama
Huntsville, Alabama

ABSTRACT. During the past decade, significant advancements in business organization modeling have been achieved. The man-organized system is a system whose elements consist of people, material, money, and information. Various methods and procedures have been developed to model this business organization. The purpose of the research project is to investigate the theoretical aspects of validating a method of modeling a man-organized system.

A new approach to modeling the organization is a method called Dynamic Organization Network Analysis (DONA). The basis for this method is the use of state variable equations. A question arises concerning the validity of the DONA model. This is answered through an experimental system designed to test the model. The experimental system incorporates a computer simulation with known characteristics. The simulation is used as a standard for comparing the performance characteristics of the DONA model.

The remainder of this article has been reproduced photographically from the authors' manuscript.

EXPERIMENTAL DESIGN CONSIDERATIONS IN

VALIDATING A METHOD OF MODELING A MAN-ORGANIZED SYSTEM

Introduction

During the past decade, significant advancements in business organization modeling have been achieved. Various procedures and methods have been developed to model systems including man-organized systems. The man-organized system is characterized by elements such as people, material, money, and information. A method widely used in modeling large organizations is computer simulation. One of the most extensive works on computer simulation of large organizations is by Jay W. Forrester.[3] In his book "Industrial Dynamics" he introduces his concept of modeling organizations and he describes a new computer language, DYNAMO, to implement the modeling technique. Forrester's method is suited to a project which requires modeling of fine details of an organization. However, the method may require years to fully model a large organization. The need for a method of modeling to be accomplished in a timely manner caused a search for other approaches.

Within the last few years, state variables have gained attention as a tool to be used in systems analysis.[1,4,6,7] The state variable equations have found use in mechanical, hydraulic and electrical systems. Herman E. Koenig, et. al. in the book "Analysis of Discrete Physical Systems"[7] proposed that the analysis may be carried into socio-economic systems. The strength of this approach lies in the fact that the systems is considered from a control-system point of view. The basic postulate is that the system is regarded as a conservative system. The matching of inputs to outputs requires formulation according to energy considerations. This approach differs from the economists approach in that the economist views the system as a black box with only inputs and outputs. Other quantities such as internal performance characteristics are not given detailed consideration by the economists. By using the control theory point of view, the analyist is capable of manipulating those quantities which control the performance of the organization.

The Dynamic Organizational Network Analysis Model

The state variable approach was used in the modeling of a socio-economic system by Koenig.[5,8] Koenig's procedure paralleled Forrester's in that mathematical relationships have to be developed for all significant operations. This is very time consuming for a large organization. The fact that only two forms of equations are needed in the state variable approach caused further considerations of this approach. The state variable equations are as follows:

$$\dot{\Psi}(t) = P\ \Psi(t) + Q\ X(t)$$
$$Y(t) = M\ \Psi(t) + N\ X(t)$$

where P, Q, M, and N are matrices characteristic of the system, $X(t)$ is the vector of inputs, $Y(t)$ is the vector of outputs, and the $\Psi(t)$ are state variables.

Rewriting these equations in matrix multiplication form yields:

$$\begin{bmatrix} \dot{\Psi}(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} P & Q \\ M & N \end{bmatrix} \cdot \begin{bmatrix} \Psi(t) \\ X(t) \end{bmatrix}$$

This form suggests a form appearing in multiple regression analysis.[2] With these concepts in mind, a modeling method called Dynamic Organizational Network Analysis (DONA) evolved.

The DONA method of modeling is a self-generating procedure in that the matrices characterizing the system are generated from the system through regression analysis. Regression analysis requires data from the system and this data will certainly consist of discrete quantities or measurements taken at some time interval. Since the state variable equations describe continuous functions, a discrete form may be derived by writing the equations in difference-equation form.

$$\begin{bmatrix} \Psi(t+h) \\ Y(t) \end{bmatrix} = \begin{bmatrix} (hP+I) & hQ \\ M & N \end{bmatrix} \cdot \begin{bmatrix} \Psi(t) \\ X(t) \end{bmatrix}$$

Taking first differences to high-pass filter the records, thus eliminating trends, yields

$$\begin{bmatrix} \Delta(\Psi) \\ \nabla(Y) \end{bmatrix} = \begin{bmatrix} (P+I) & Q \\ M & N \end{bmatrix} \cdot \begin{bmatrix} \nabla(\Psi) \\ \nabla(X) \end{bmatrix} \quad \text{for } h = 1$$

where

$$\Delta(\Psi) = \Psi(t+1) - \Psi(t)$$

$$\nabla(\Psi) = \Psi(t) - \Psi(t-1)$$

Letting the matrix $[S] = \begin{bmatrix} (P+I) & Q \\ M & N \end{bmatrix}$ and substituting yields

$$\begin{bmatrix} \Delta(\Psi) \\ \nabla(Y) \end{bmatrix} = \begin{bmatrix} S \end{bmatrix} \cdot \begin{bmatrix} \nabla(\Psi) \\ \nabla(X) \end{bmatrix}$$

The ∇X, ∇Y, $\nabla \Psi$, and $\Delta \Psi$ are determined from calculations using actual data produced by the real system to be modeled. The time interval may be one hour, one week, etc. By the use of a computer program for multiple regression analysis, the characteristic matrix, S, can be determined.

The Laboratory Concept for Validation

A question arises concerning the validity of the DONA model and the method which produced the model. The usual procedure for validating such a model is by using information from the real system and determining if the model predicts in an acceptable manner when compared with the performance of the real system. It was felt that a better method could be used to validate the DONA model. A laboratory concept was developed to validate the method, which in turn validates the model produced by the method. In this concept the paramaters can be controlled to determine the range and responsiveness of the model. The experimental system shown in Figure 1 is used for the validation procedure. The block marked "SIMCO" is a computer simulation of a sales company. This simulation was developed by C. McMillan and Richard F. Gonzales.[9] As orginally written SIMCO was a distributor operation for a single product. Stochastic demand and lead times were incorporated. SIMCO was modified to handle a second product and to simulate personnel actions; i.e. hires, fires, and transfers. These modifications were made to widen the scope of operations through the addition of the second product and to have some interaction of elements; e.g. the transfer of personnel from one product line to the other. All of the characteristics of SIMCO are known. Since SIMCO is a subroutine, it can easily be replaced to study the validity of DONA as applied to any other simulated activity.

The DONA methodology block in Figure 1 contains the multiple regression analysis procedure for developing the "S" matrix. In the laboratory system the "S" matrix will characterize the SIMCO Sales Company. The DONA model block in Figure 1 represents the following matrix equation.

$$\begin{bmatrix} \Psi(t+1) \\ Y(t) \end{bmatrix} = \begin{bmatrix} S \end{bmatrix} \cdot \begin{bmatrix} \Psi(t) \\ X(t) \end{bmatrix}$$

The left side of the matrix equation is the DONA output.

The outputs of SIMCO and the DONA model are finally compared as shown in Figure 1. The comparison is made on all parameters desired or deemed appropriate for consideration. For example, the DONA model not only produces the system outputs but also predicts for the next time interval the value of the state variables.

Experimental Design Considerations in the Validation Procedure

The problems revealed in this project have provided some valuable insights into modeling of a man-organized system using this method. The laboratory concept described above has been fully implemented. The entire concept has been written into a computer program and the program has been run and debugged.

Since this is a laboratory, the generation of stimulus data was the first big problem. Even though the computer program could handle a total of 40 input and state variables the problem was not the number of variables. The problems centered around the behavior of the variables.

It was learned that a man-organized system as simulated produces variables some of which cause redundancies in the equations. This of course causes singularities in the matrix. A particular variable of the stimulus data, though time-varying over the long term, may have a constant value for the simulated time period; and when the forward and backward differences are taken, the value of the difference is zero. This same problem can occur with a parameter whose first derivative is a constant. The forward and backward differences will be constant but in the multiple regression analysis, a zero variance is computed. A possible solution to the problems of zero values for the variances is the use of a "dithering signal," much the same as the dithering signal used in control systems. The redundancies in the equations can be overcome by careful selection of the system parameters.

After taking careful note of the above conditions, attention is then directed to the types and levels of stimulus data. The prediction capabilities of the DONA model can be tested through its ability to "track" the real system or, as in the case of the laboratory concept, to "track" the SIMCO simulation. The inputs to SIMCO and DONA may be any one or a combination of signals composed of random noise, sinusoids, impluses or step functions. A particular characteristic to observe is the frequency response of the DONA model. The interactions designed into the SIMCO simulation were part of this test to determine the "worth" of the DONA model and the method to produce the model.

Problems can also arise in the comparison phase of the laboratory concept. The criterion for agreement in the outputs is based on the quadratic form of the covariance matrix. A recent book by Jenkins and Watts[10] describes this procedure when an analysis is required of a multivariate system. The equation of the quadratic form of the covariance matrix is as follows:

$$\text{Probability } \{[\vec{X} - \vec{\mu}]^T \; V^{-1}[\vec{X} - \vec{\mu}]\} = \text{Probability } (\chi^2{}_m)$$

$$\text{where } V = [\text{cov}_T \{y(j,t),\ y(k,t)\}] \;, \; t = 1 \ldots, T$$

$$m = \text{degrees of freedom} = \text{order of } V \text{ matrix}$$

This equation is used in two different statistical tests. First, the equation is used to test the prediction capability of the model as a function of time. It can be determined when in the time domain the model ceases to predict with the specified confidence. The equation then may be used to test the significance of each dimension (variable) in the model. Instead of measuring the variability as a function of time the variability is measured as a function of a variable in the model with the time fixed. Again it can be observed when the model ceases to predict with the specified confidence. Using this method for comparison of the outputs quantitative information is generated which gives the performance characteristics of the DONA model and the method used to produce the DONA model.

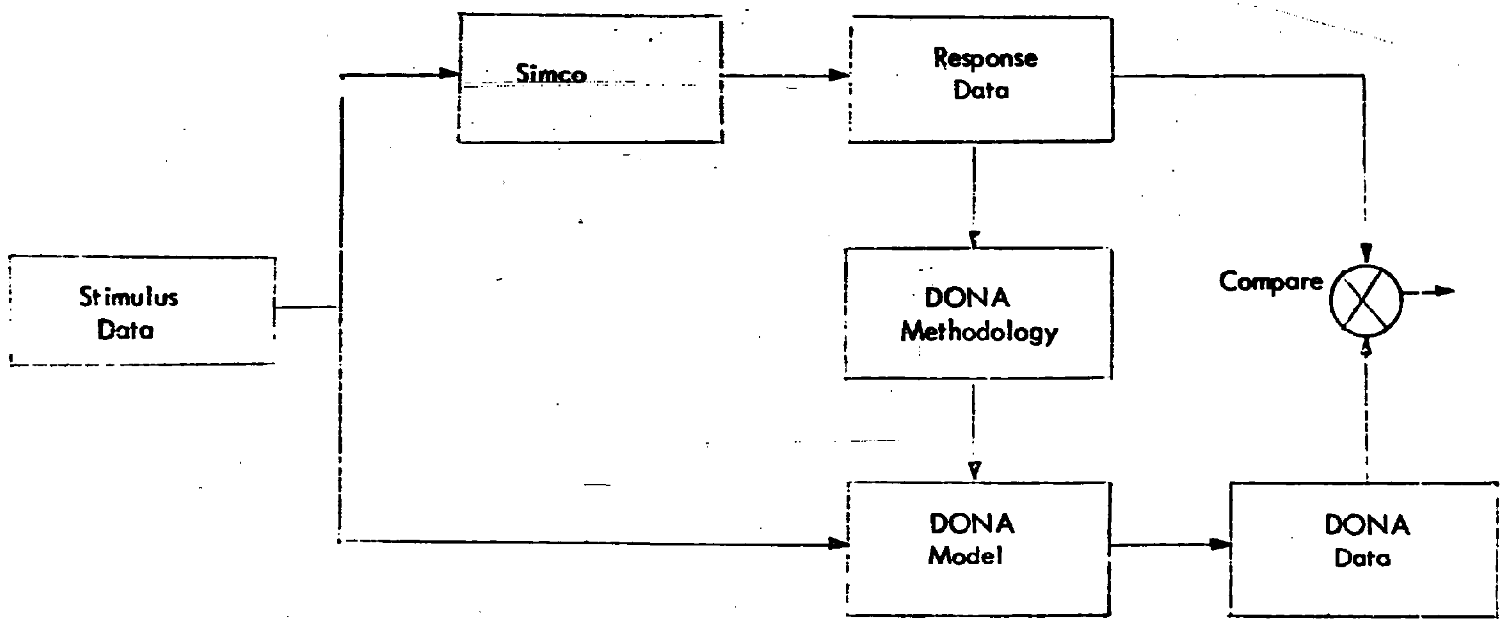


Figure 1 The Laboratory System

BIBLIOGRAPHY

1. Bashkow, Theodore, R., "The A Matrix, New Network Description." IRE Transactions on Circuit Theory, Vol. CT-4, No. 3, (September, 1957), pp. 117-119

2. Duncan, Acheson J., Quality Control and Industrial Statistics. Homewood, Ill., Richard D. Irwin, Inc., 1965.

3. Forrester, Jay W., Industrial Dynamics. New York, the M.I.T. Press and John Wiley, 1961.

4. Frame, J. S. and Koenig, H. E., "Matrix Functions and Applications. Part III-Applications of Matrices to Systems Analysis." IEEE Spectrum, May, 1964, pp. 100-109.

5. Koenig, H. E., "Mathematical Models of Socio-Economic Systems: An Example." IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-1, No. 1 (Nov., 1965), pp. 41-45.

6. Koenig, Herman E., Tokad, Yilmaz, and Kesavan, Analysis of Discrete Physical Systems, New York, McGraw-Hill, 1967.

7. Koenig, H. E., and Tokad, Y., "State Models of Systems of Multi-Terminal Linear Components." IEEE 1964 International Convention Record, Part I, Vol. 12, pp. 318-329.

8. Koenig, H. E., Kenney, M. G., and Zenach R., A System Model for Management, Planning, and Resource Allocation in Institutions of Higher Education. Final Report to the National Science Foundation, Washington, D. C. Project C-518. September 30, 1968.

9. McMillan, Claude, and Gonzalez, Richard F., Systems Analysis: A Computer Approach to Decision Models. Homewood, Ill., Richard D. Irwin, Inc., 1965.

10. Jenkins, Gwilym M., and Watts, Donald G., Spectral Analysis and Its Applications. San Francisco, Calif., Holden-Day, 1969.

An Investigation of the Effect of Some Prior Distributions on Bayesian Confidence Intervals For Attribute Data

Alan W. Benton
Aberdeen Research and Development Center
Aberdeen Proving Ground, Maryland

ABSTRACT

One method of obtaining confidence intervals on the reliability of a system or component whose sample outcomes are either a success or failure is founded on Bayes theorem. In the Bayesian formulation of the problem one must assume a prior distribution on the random variable of interest, namely, the reliability. The purpose of this investigation was to determine the effect of some prior distributions on Bayesian confidence intervals in which it was assumed that the prior distribution may be represented by the beta distribution. It was my intent to restrict attention to alternative priors one might use when no previous data or experience exists on a system.

INTRODUCTION

Of primary importance to Bayesian statistics is the attachment of probabilities to various possible hypotheses, or in this case probable reliabilities. The mechanism that performs this role is the prior distribution. The plausible values which the random variable might take on do not necessarily imply that one believes the probability exactly, they are only a measure or rough indication of what one tends to believe are the most likely values.

In Bayesian analysis the prior distribution is combined with the test data to yield a modified distribution of reliability, namely, the posterior distribution. That is, the information that we have on a component by way of the prior distribution is updated with the latest test results. If we let

This article has been reproduced photographically from the author's manuscript.

W = Reliability, $0 \leq W \leq 1$

n = Number of tests

r = Number of successes

then by Bayes theorem

$$f(W/n,r) = \frac{g(n,r/W)f(W)}{\int_W g(n,r/W)f(W)dW}$$

where

$f(W)$ = Prior distribution of W

$g(n,r/W)$ = Probability of observing r successes in n trials, given W.

$f(W/n,r)$ = Posterior distribution of W.

In practice it is quite common to assume a beta for the prior distribution. There are several reasons for suggesting the beta, among them are (1) its random variable has range [0, 1], (2) the beta can fit almost any unimodal or nonmodal distribution for a r.v. over the unit interval, and (3) its ease in computations. The beta density is given by

$$f(W) = \frac{(a+b+1)!}{a!\,b!} W^a(1-W)^b, \quad 0 \leq W \leq 1$$
$$= 0 \qquad \text{elsewhere}$$

where $a,b > -1$. It may be shown that by altering a and b the shape of the distribution is changed. See Figure 1. These distributions indicate the plausible values which the random variable might take on.

Assuming $f(W)$ to be beta distributed and $g(n,r/W)$ to be a binomial distribution, the posterior distribution is given by

$$f(W/n,r) = \frac{g(n,r/W)f(W)}{\int_W g(n,r/W)f(W)dW}$$

$$= \frac{\binom{n}{r} W^r (1-W)^{n-r} \frac{(a+b+1)!}{a!b!} W^a(1-W)^b}{\int_0^1 \binom{n}{r} W^r(1-W)^{n-r} \frac{(a+b+1)!}{a!b!} W^a(1-W)^b dW}$$

$$= \frac{(a+b+n+1)!}{(a+r)!(b+n-r)!} W^{a+r} (1-W)^{b+n-r}$$

which is also beta distributed.

CONSTRUCTION OF CONFIDENCE INTERVALS

Since the posterior distribution was found to be beta distributed, it follows that a confidence interval for a beta variate is desired. If we let $W \sim B(a,b)$, then a $100(1-\alpha)\%$ one sided confidence interval is given by

$$\Pr\left[W_L^1 < W < 1\right] = \int_{W_L^1}^{1} f(W)\,dW = 1-\alpha$$

An exact lower bound is given by

$$\frac{1}{1 + \frac{b+n-r+1}{a+r+1} F_{1-\alpha}}$$

where

$F_{1-\alpha}$ = $(1-\alpha)$ percentile of the F- distribution with $2(b+n-r+1)$ and $2(a+r+1)$ degrees of freedom.

a,b = parameters from the prior

n = sample size

r = number of successes

The use of the F- distribution may be seen from the following. From the change of variable theorem of integral calculus we may write

$$g(F) = h(W = U(F)) \frac{dW}{dF}.$$

Let

$$W = \frac{\frac{a+r+1}{b+n-r+1} F}{1 + \frac{a+r+1}{b+n-r+1} F}$$

then

$$\frac{dW}{dF} = \frac{\frac{a+r+1}{b+n-r+1}}{\left(1 + \frac{a+r+1}{b+n-r+1} F\right)^2}$$

and on substitution

$$g(F) = \frac{\Gamma(a+b+n+2)}{\Gamma(a+r+1)\,\Gamma(b+n-r+1)} \left(\frac{a+r+1}{b+n-r+1}\right)^{a+r+1} \frac{F^{a+r}}{\left(1+\frac{a+r+1}{b+n-r+1}F\right)^{a+b+n+2}}$$

which is the F- distribution with 2(a+r+1) and 2(b+n-r+1) degrees of freedom.

Now recall

$$\int_{W_L^1}^{1} f(W)\,dW = \int_{\frac{RF}{1+RF}} f(W)\,dW = 1-\alpha$$

where

$$R = \frac{a+r+1}{b+n-r+1}$$

Noting that $F_b^a(\alpha) = 1/F_a^b(1-\alpha)$, the lower limit is given by

$$W_L^1 = \frac{1}{1+\frac{b+n-r+1}{a+r+1}F_{1-\alpha}}$$

One of the reasons for using the F- distribution is that tables of this distribution are frequently more readily available than those of the incomplete beta. Also, they are convenient for those values of α most often used, e.g., 0.10, 0.05, 0.01.

For sufficiently large sample sizes the normal approximation may be employed, the approximation being best in the vicinity of W equal to one-half. For the posterior distribution the expected value and variance of W are:

$$E(W) = \frac{a+r+1}{b+n-r+1}$$

and

$$\sigma^2 = \frac{(a+r+1)(b+n-r+1)}{(a+b+n+2)^2(a+b+n+3)}$$

Thus, the 100(1-α)% lower limit is given by $E(W) - z_{1-\alpha}\,\sigma$

where $z_{1-\alpha}$ is such that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_{1-\alpha}} e^{-t^2/2}\, dt = 1-\alpha.$$

DISCUSSTION

As indicated in the previous section, the lower confidence limit (LCL) was computed for illustrative purposes. A confidence level of 95% was chosen and the following prior distributions were selected:

a	-1	-1/2	0	1	6
b	-1	-1/2	0	0	1

References [1] and [4] give discussion on the reasons for assuming prior distributions when a and b are both set equal to -1, -1/2, and 0. By letting a=6 and b=1 we give more credence to moderately high reliabilities, which would appear to be a likely area of interest. For large a and/or b one gets into the problem of the prior distribution far outweighing the most recent evidence or sample.

In Figure 2 the lower 95% limit is plotted against the number of failures observed in a sample of size twenty. In general, the results are quite similar for the other sample sizes investigated (n=10(5)25). It may be noted that the number of failures begins at one. This was done since when using the (-1, -1) prior one must assume the occurrence of at least one success and at least one failure in the sample.

An examination of the lower limits plotted on Figure 2 yields the following general results:

1. The (-1, -1) prior results in shorter confidence intervals than a(-1/2, -1/2) prior which in turn gives a shorter confidence interval than the (0,0) prior. This holds true for moderate to high reliabilities. The

order is reversed for low reliabilities.

2. The differences in confidence interval length become smaller as n, the sample size, increases.

3. All three priors [(-1,-1), (-1/2, -1/2), (0,0)] result in shorter confidence intervals than those obtained with classical methods.

4. The (6,1) prior does not result in a uniformly shorter confidence interval for moderate to high reliabilities.

Now, let r/n (# success/sample) be an indicator of reliability and rank the priors according to their interval length. Thus the shortest confidence interval indicates the most optimistic result and the longest the most pessiminstic result for sample estimates of reliability. The following table presents a summary of these rankings.

Summary of Length of Confidence Intervals for

(-1, -1), (-1/2, -1/2), (0, 0), (1, 0) Priors

r/n	Shortest	——	——	Longest
$\geqslant 0.85$	(-1, -1)	(-1/2, -1/2)	(1,0)	(0,0)
0.75-0.85	(-1, -1)	(1, 0)	(1/2, - 1/2)	(0, 0)
0.65-0.75	(1, 0)	(-1, -1)	(-1/2, -1/2)	(0, 0)

Thus, for r/n greater than 0.75 the (-1, -1) prior gives the shortest interval while the uniform prior (o, 0) gives the longest interval.

References

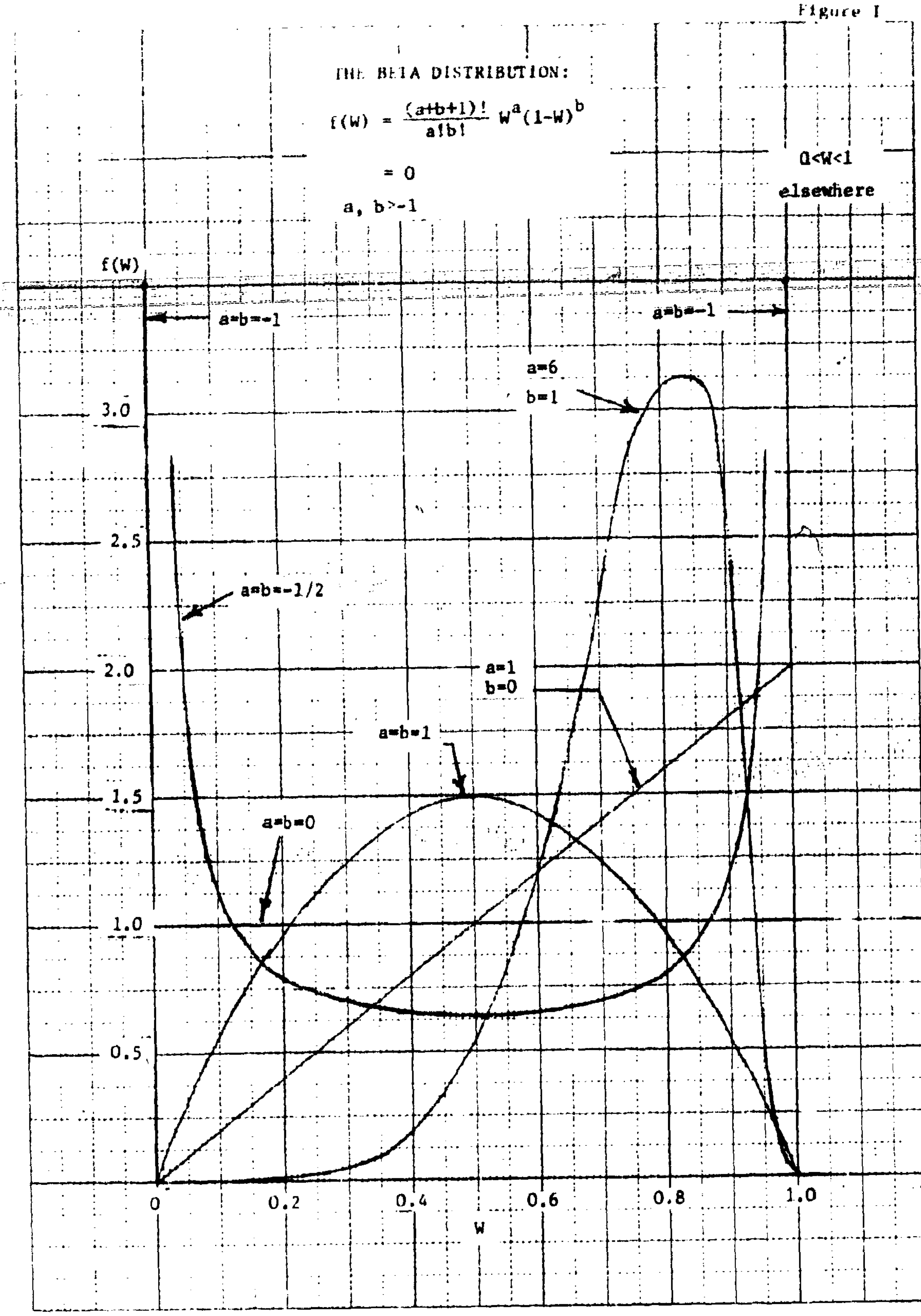
1. Lindley, D. V., Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2, Cambridge University Press, New York, 1965.

2. Mood, A. M., Graybill, F. A. Introduction to the Theory of Statistics, McGraw-Hill Book Co., Inc., New York, 1963.

3. Savage, L. J., The Foundations of Statistics, John Wiley and Sons, Inc., New York, 1954.

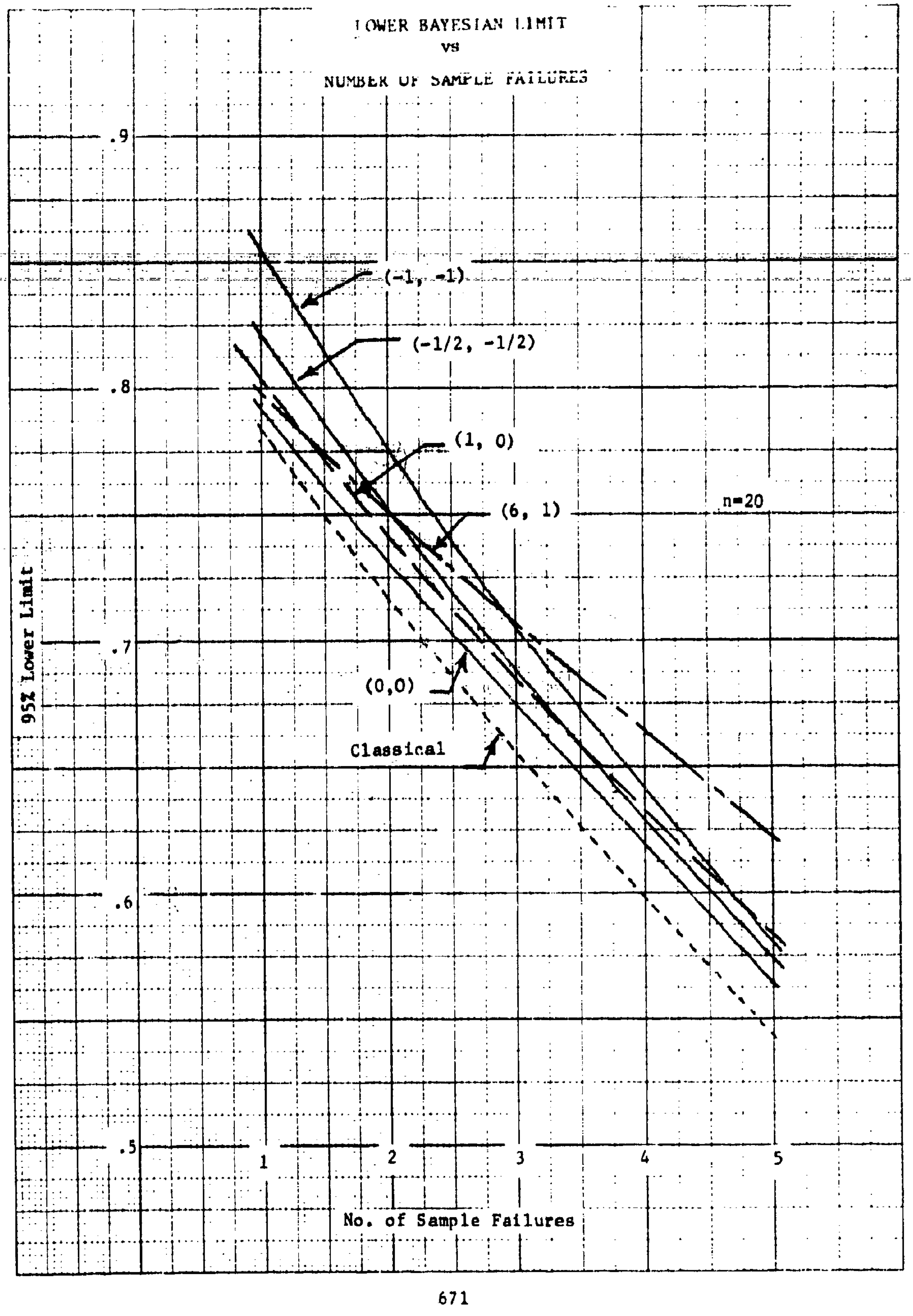
4. "Planning Guide for Demonstration and Assessment of Reliability and Durability", AMC Pamphlet AMCP 702-5.

Figure I



THE BETA DISTRIBUTION:
f(W) = (a+b+1)! / a!b! W^a (1-W)^b   0<W<1
= 0   elsewhere
a, b>-1
f(W)
a=b=-1
a=b=-1
a=6
b=1
3.0
2.5
a=b=-1/2
2.0
a=1
b=0
a=b=1
1.5
a=b=0
1.0
0.5
0
0.2
0.4
0.6
0.8
1.0
W

Figure 2


LOWER BAYESIAN LIMIT
vs
NUMBER OF SAMPLE FAILURES
.9
.8
.7
.6
.5
95% Lower Limit
(-1, -1)
(-1/2, -1/2)
(1, 0)
(6, 1)
n=20
(0,0)
Classical
1
2
3
4
5
No. of Sample Failures

SOME TECHNIQUES FOR CONSTRUCTING
MUTUALLY ORTHOGONAL LATIN SQUARES*

W. T. Federer, A. Hedayat, E. T. Parker
B. L. Raktoe, Esther Seiden, and R. J. Turyn
Mathematics Research Center, U. S. Army
Madison, Wisconsin

ABSTRACT. Various methods of constructing a set of mutually orthogonal latin squares are presented and the theoretical aspects of various methods are discussed. Illustrative examples of constructing latin squares and sets of mutually orthogonal latin squares are given. The methods of constructing latin squares and sets of orthogonal latin squares are complete and partial confounding, fractional replication, analysis of variance, group, projecting diagonals, orthomorphism, pairwise balanced design, oval, code, product composition, and sum composition. The methods of construction designated as partial confounding, fractional replication, analysis of variance, and sum composition appear not to have been discussed previously in the literature. The methods of complete confounding and of projecting diagonals have been given only a passing reference with no indication as to the actual construction procedure. The sum composition method has interesting consequences in combinatorial theory as well as in the construction of orthogonal latin squares. Lastly, equivalences of fourteen combinatorial systems to orthogonality in latin squares has been investigated and described.

*This article has appeared as Technical Summary Report No. 1030 of The Mathematics Research Center, U. S. Army, The University of Wisconsin, Madison, sponsored under Contract No.: DA-31-ARO-D-462.
The remainder of this article has been reproduced photographically from the authors' manuscript.

TABLE OF CONTENTS

SOME TECHNIQUES FOR CONSTRUCTING MUTUALLY ORTHOGONAL LATIN SQUARES

W. T. Federer[1], A. Hedayat[2], E. T. Parker[3]
B. L. Raktoe[4], Esther Seiden[5], and R. J. Turyn[6]

I. Introduction and Some Terminology

The purpose of this paper is to present a set of methods for constructing mutually orthogonal latin squares and to exhibit some squares produced by each of the methods. The set of methods presented herein was discussed in a series of informal seminars held during the weeks of July 14-18 and 21-25, 1969, by the authors at Cornell University. The motivation for these discussions was derived from results obtained by Hedayat [1969] and from the optimism of the authors. New procedures for constructing a set of mutually orthogonal latin squares and new views of present methods of construction were desired in order to advance the theory of mutual orthogonality in latin squares.

---

[1] Professor of Biological Statistics, Cornell University and Visiting Professor, Mathematics Research Center, U.S. Army and University of Wisconsin (on sabbatical leave 1969-70).

[2] Assistant Professor, Cornell University.

[3] Professor Mathematics, University of Illinois, and Visiting Professor, Cornell University (July, 1969).

[4] Associate Professor, University of Guelph and Visiting Associate Professor, Cornell University (January to August, 1969).

[5] Professor, Michigan State University, and Visiting Professor, Cornell University (June, July, August, 1969).

[6] Mathematician, Raytheon Corporation, and Visiting Professor, Cornell University (July, 1969).

As may be noted from the table of contents, the different sections were written by different authors. An attempt was made to have a consistent notation and a uniform style. Although much more work is required to finalize the method in several of the sections enough is known about the method to use it to construct a latin square of any order or to construct a set of two or more mutually orthogonal latin squares. Also, a number of equivalences may be noted for some of the methods.

The theory of mutual orthogonality in latin squares has application in the construction of many classes of experiment designs and in many combinatorial systems. The latter subject is discussed in section XV where the equivalences of various combinatorial systems are presented. With regard to the former subject, there is an ever present need for new experiment designs for new experimental situations in order for the experimenter not to have to conduct his experiment to fit known experiment designs.

Some of the notation and terminology that will be utilized is presented below.

Definition 1.1. A latin square of order n on a set $\sum$ with n distinct elements is an $n \times n$ matrix each of whose rows and columns is a permutation of the set $\sum$.

Example:

$$\begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 2 & 3 & 1 \\ \hline 3 & 1 & 2 \\ \hline \end{array}$$ is a latin square of order 3 on $\sum = \{1, 2, 3\}$.

Definition 1.2. Two latin squares $L_1 = (a_{ij})$ and $L_2 = (b_{ij})$ of order n are said to be orthogonal if the n^2 ordered pairs (a_{ij}, b_{ij}) $(i,j = 1,2,\ldots,n)$ are all distinct. Note that L_1 and L_2 need not be defined on the same set.

Example:

1	2	3
2	3	1
3	1	2

and

A	B	C
C	A	B
B	C	A

.

Definition 1.3. The members of a set of t latin squares $L_1, L_2, \ldots, L_t$ of order n are said to be mutually (pairwise) orthogonal if L_i is orthogonal to L_j, $i \neq j$, $i,j = 1,2,\ldots,t$. Hereafter by an $O(n,t)$ set we mean a set consisting of t mutually orthogonal latin squares of order n.

Example:

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

,

1	2	3	4
4	3	2	1
2	1	4	3
3	4	1	2

, and

1	2	3	4
3	4	1	2
4	3	2	1
2	1	4	3

.

Latin squares and orthogonal latin squares have at least 187 years of history. Hedayat [1969-section IX] has presented a reasonably good picture of this history which will not be repeated here. It is planned to prepare a historical account of developments related to orthogonality in latin squares and to publish this material together with a bibliography elsewhere.

II. Factorial Confounding Construction of $O(n,t)$ Sets

II.1. Complete Confounding

The n^2 row-column intersections may be related to the treatment combinations in a n^2 factorial treatment design. To illustrate let us consider the 4^2 factorial and the latin square of order 4. The levels of the main effects, A and B, in the factorial will be used to designate the rows and the columns of the latin square of order 4 as follows:

latin squares of order 4

	column 1 = $(B)_0$	column 2 = $(B)_1$	column 3 = $(B)_2$	column 4 = $(B)_3$
row 1 = $(A)_0$	00	01	02	03
row 2 = $(A)_1$	10	11	12	13
row 3 = $(A)_2$	20	21	22	23
row 4 = $(A)_3$	30	31	32	33

Thus, four combinations out of the 16 which have $i = 0$ in the subscript ij, i.e., 00, 01, 02, and 03, are designated as $(A)_0$ and are put in row 1. Continuing this procedure the remainder of the ij combinations are allocated to the remaining rows and to the columns as shown above.

Now, three other effects with 4 levels each can be set up from the projective geometry $PG(1,2^2)$; the effects are $(AB^{u_1})_{u_i+u_1u_j}$, $(AB^{u_2})_{u_i+u_2u_j}$, and $(AB^{u_3})_{u_i+u_3u_j}$. The levels of these effects and the corresponding latin square produced by letting all combinations of the level of an effect be a symbol in the latin square are (see page 337 of Kempthorne [1952], e.g.):

$$(AB^{u_1})_{u_i+u_1u_j} = \begin{cases} 0 & 00 + 11 + 22 + 33 \to I \\ 1 & 01 + 10 + 23 + 32 \to II \\ 2 & 02 + 13 + 20 + 31 \to III \\ 3 & 03 + 12 + 21 + 30 \to IV \end{cases}$$

00 = I	01 = II	02 = III	03 = IV
10 = II	11 = I	12 = IV	13 = III
20 = III	21 = IV	22 = I	23 = II
30 = IV	31 = III	32 = II	33 = I

$$(AB^{u_2})_{u_i+u_2u_j} = \begin{cases} 0 & 00 + 13 + 21 + 32 \to \alpha \\ 1 & 03 + 10 + 22 + 31 \to \beta \\ 2 & 01 + 12 + 20 + 33 \to \gamma \\ 3 & 02 + 11 + 23 + 30 \to \delta \end{cases}$$

α	γ	δ	β
β	δ	γ	α
γ	α	β	δ
δ	β	α	γ

$$(AB^{u_3})_{u_i+u_3u_j} = \begin{cases} 0 & 00 + 12 + 23 + 31 \to W \\ 1 & 02 + 10 + 21 + 33 \to X \\ 2 & 03 + 11 + 20 + 32 \to Y \\ 3 & 01 + 13 + 22 + 30 \to Z \end{cases}$$

W	Z	X	Y
X	Y	W	Z
Y	X	Z	W
Z	W	Y	X

In the above the complete confounding scheme of sources of variation in the O(4,3) set and the effects in the factorial may be illustrated in the following analysis of variance table wherein the sums of squares in the lines of the analysis of variance are orthogonal to each other:

Source of variation	Degrees of freedom
Correction for mean	1
Rows = A effect	3
Columns = B effect	3
Roman numbers = (AB^{u_1}) effect	3
Greek letters = (AB^{u_2}) effect	3
Latin letters = (AB^{u_3}) effect	3
Total	16

Instead of relating the mutually orthogonal latin squares of order 4 to a 4^2 factorial we may relate them to a 2^4 factorial in the following manner. Let the 16 row-column intersections be numbered as follows:

	column			
row	1	2	3	4
1	0000	0001	0010	0011
2	0100	0101	0110	0111
3	1000	1001	1010	1011
4	1100	1101	1110	1111

Let the factors be a, b, c, and d with two levels (0 and 1) each. The rows correspond to factorial effects A, B, and AB and the columns correspond to factorial effects C, D, and CD. (This form of constructing latin squares has been used by Fisher and Yates [1957] for latin squares of order 8.) Then, let the symbols in the 3 latin squares be represented by the following scheme:

Factorial generators	Combinations	latin squares			
$(AC)_0$, $(BD)_0$, $(ABCD)_0$	0000 + 0101 + 1010 + 1111 = I	I	II	III	IV
$(AC)_0$, $(BD)_1$, $(ABCD)_1$	0001 + 0100 + 1011 + 1110 = II	II	I	IV	III
$(AC)_1$, $(BD)_0$, $(ABCD)_1$	0010 + 0111 + 1000 + 1101 = III	III	IV	I	II
$(AC)_1$, $(BD)_1$, $(ABCD)_0$	0011 + 0110 + 1001 + 1100 = IV	IV	III	II	I
$(AD)_0$, $(ABC)_0$, $(BCD)_0$	0000 + 0110 + 1011 + 1101 = W	W	Z	X	Y
$(AD)_0$, $(ABC)_1$, $(BCD)_1$	0010 + 0100 + 1001 + 1111 = X	X	Y	W	Z
$(AD)_1$, $(ABC)_0$, $(BCD)_1$	0001 + 0111 + 1010 + 1100 = Z	Y	X	Z	W
$(AD)_1$, $(ABC)_1$, $(BCD)_0$	0101 + 0011 + 1000 + 1110 = Y	Z	W	Y	X

$(ACD)_0$, $(BC)_0$, $(ABD)_0$ $\quad 0000 + 0111 + 1110 + 1001 = \alpha$

$(ACD)_0$, $(BC)_1$, $(ABD)_1$ $\quad 1010 + 0100 + 0011 + 1101 = \beta$

$(ACD)_1$, $(BC)_0$, $(ABD)_1$ $\quad 1000 + 0110 + 1111 + 0001 = \gamma$

$(ACD)_1$, $(BC)_1$, $(ABD)_0$ $\quad 0010 + 0101 + 1011 + 1100 = \delta$

α	γ	δ	β
β	δ	γ	α
γ	α	β	δ
δ	β	α	γ

The correspondence of the latin squares obtained from complete confounding considering a 4^2 factorial and considering a 2^4 factorial is demonstrated in the following analysis of variance table:

Source of variation		degrees of freedom	
Correction for mean		1	
Rows = A effect in 4^2 factorial		3	
	A effect in 2^4 factorial		1
	B " " 2^4 "		1
	AB " " 2^4 "		1
Columns = B effect in 4^2 factorial		3	
	C effect in 2^4 factorial		1
	D " " 2^4 "		1
	CD " " 2^4 "		1
Roman numbers = AB^{u_1} effect in 4^2 factorial		3	
	AC effect in 2^4 factorial		1
	BD " " 2^4 "		1
	ABCD " " 2^4 "		1
Greek letters = AB^{u_2} effect in 4^2 factorial		3	
	ACD effect in 2^4 factorial		1
	BC " " 2^4 "		1
	ABD " " 2^4 "		1

Latin letters = AB^{u} effect in 4^2 factorial	3	
AD effect in 2^4 factorial		1
ABC " " 2^4 "		1
BCD " " 2^4 "		1
Total		16

It should be noted here that the effects in the 2^4 map directly into the 4^2 projective geometry or $PG(1, 2^2)$. Likewise, even though one more set of generators is available, viz.

	Generators	Interaction
Roman numbers	= AD, BC	ABCD
Greek letters	= AC, ABD	BCD
Latin letters	= BD, ABC	ACD

the three orthogonal latin squares produced are the same ones. Since the third effect above is obtained as the product of the two generators, mod 2, we need consider only the generators. Multiplying these by CD (mod 2) we obtain the generators of the preceding scheme. Hence, even though two different complete confounding schemes are available there is a simple one-to-one mapping of one set into the other set. Although nothing interesting turns up here, it would be interesting to study the various complete confounding schemes in the latin square of order 9 as related to the 3^4 factorial.

As a second illustration of the use of complete confounding to construct latin squares, let us consider the latin square of order 6. Using the notation and concepts of Raktoe [1969] we designate the 6^2 as a $2^2(3)^2$ factorial and represent a combination by ghij where g, h are members of I (3) and i, j are members of I (4). The effects in the 2^2 and in the 3^2 factorials are denoted respectively by:

$$\begin{array}{ll} A^3 & C^4 \\ B^3 & D^4 \\ A^3B^3 & C^4D^4 \\ & C^4D^2 \end{array}$$

The remaining interactions are given below in the analysis of variance table:

Source of variation	Degrees of freedom	
Correction for mean	1	
Rows = A^3C^4	5	
A^3		1
C^4		2
$A^3 \times C^4$		2
Columns = B^3D^4	5	
B^3		1
D^4		2
$B^3 \times D^4$		2
Treatments or symbols = $A^3B^3C^4D^4$	5	
A^3B^3		1
C^4D^4		2
$A^3B^3 \times C^4D^4$		2

Remainder	20	
C^4D^2		2
$A^3 \times D^4$		2
$A^3 \times C^4D^4$		2
$A^3 \times C^4D^2$		2
$B^3 \times C^4$		2
$B^3 \times C^4D^4$		2
$B^3 \times C^4D^2$		2
$A^3B^3 \times C^4$		2
$A^3B^3 \times D^4$		2
$A^3B^3 \times C^4D^2$		2
Total	36	

Let us now set up the 6 rows and the 6 columns of a latin square of order 6 with the corresponding designation of the 36 combinations as follows:

Columns

Rows	$(B^3D^4)_0$	$(B^3D^4)_1$	$(B^3D^4)_2$	$(B^3D^4)_3$	$(B^3D^4)_4$	$(B^3D^4)_5$
$(A^3C^4)_0$	0000	0304	0002	0300	0004	0302
$(A^3C^4)_1$	3040	3344	3042	3340	3044	3342
$(A^3C^4)_2$	0020	0324	0022	0320	0024	0322
$(A^3C^4)_3$	3000	3304	3002	3300	3004	3302
$(A^3C^4)_4$	0040	0344	0042	0340	0044	0342
$(A^3C^4)_5$	3020	3324	3022	3320	3024	3322

Now let the levels of $A^3B^3C^4D^4$ correspond to the symbols in a latin square of order 6 as follows:

Levels	Combination for which 3g+3h+4i+4j, mod 6, is constant		Symbol
$(A^3B^3C^4D^4)_0$	0000 + 3342 + 0024 + 3300 + 0042 + 3324	→	0
$(A^3B^3C^4D^4)_1$	0304 + 3040 + 0322 + 3004 + 0340 + 3022	→	1
$(A^3B^3C^4D^4)_2$	0002 + 3344 + 0020 + 3302 + 0044 + 3320	→	2
$(A^3B^3C^4D^4)_3$	0300 + 3042 + 0324 + 3000 + 0342 + 3024	→	3
$(A^3B^3C^4D^4)_4$	0004 + 3340 + 0022 + 3304 + 0040 + 3322	→	4
$(A^3B^3C^4D^4)_5$	0302 + 3044 + 0320 + 3002 + 0344 + 3020	→	5

This produces the following latin square of order 6:

0	1	2	3	4	5
1	2	3	4	5	0
2	3	4	5	0	1
3	4	5	0	1	2
4	5	0	1	2	3
5	0	1	2	3	4

Alternatively we could have used levels of $A^3B^3C^4D^2$ to construct the following latin square of order 6:

Levels	Combinations for which 3g+3h+4i+2j, mod 6, is constant	Symbol
$(A^3B^3C^4D^2)_0$	0000 + 3344 + 0022 + 3300 + 0044 + 3322	0
$(A^3B^3C^4D^2)_1$	0302 + 3040 + 0324 + 3002 + 0340 + 3024	1
$(A^3B^3C^4D^2)_2$	0004 + 3342 + 0020 + 3304 + 0042 + 3320	2
$(A^3B^3C^4D^2)_3$	0300 + 3044 + 0322 + 3000 + 0344 + 3022	3
$(A^3B^3C^4D^2)_4$	0002 + 3340 + 0024 + 3302 + 0040 + 3324	4
$(A^3B^3C^4D^2)_5$	0304 + 3042 + 0320 + 3004 + 0342 + 3020	5

latin square of order 6

0	5	4	3	2	1
1	0	5	4	3	2
2	1	0	5	4	3
3	2	1	0	5	4
4	3	2	1	0	5
5	4	3	2	1	0

Thus, the above square is simply a column permutation of the previous one. As there are no other sets of 5 degrees of freedom leading to a latin square of order 6 (i.e. A^3, B^3, and A^3B^3 exhaust the three single degrees of freedom from the 2^2 factorial and C^4, D^4, C^4D^4, and C^4D^2 exhaust all sets of 2 degrees of freedom from the 3^2 factorial), it is not possible to obtain a latin square of order 6 orthogonal to either of the preceding ones using complete confounding schemes.

For a latin square of order 10 we may use levels of $A^5B^5C^6D^6$, $A^5B^5C^6D^2$, $A^5B^5C^6D^8$, or $A^5B^5C^6D^4$ to form four different latin squares of order 10.

II. 2. Partial Confounding

In the last section use was made of complete confounding of effects in a factorial with the rows, columns, and symbols in a latin square. Instead of completely confounding an effect, it could be partially confounded. For example, the latin square of order 4 could be considered as a 2^4 factorial as in the preceding section, with the following scheme of confounding:

Rows	$1 = (C)_0$	$2 = (C)_1$	$3 = (D)_0$	$4 = (D)_1$
1 $(A)_0, (B)_0$	0000	0011	0010	0001
2 $(A)_0, (B)_1$	0101	0110	0100	0111
3 $(A)_1, (B)_0$	1000	1011	1010	1001
4 $(A)_1, (B)_1$	1101	1110	1100	1111

If we set up the latin square symbols for the above as

α	β	γ	δ
β	α	δ	γ
γ	δ	β	α
δ	γ	α	β

then

the symbols correspond to the following combinations:

α: $0000 + 0110 + 1001 + 1100 = (ABCD)_0$ + other effects

β: $0011 + 0101 + 1010 + 1111 = (ABCD)_0$ + " "

γ: $1000 + 1110 + 0010 + 0111 = (ABCD)_1$ + " "

δ: $0001 + 0100 + 1011 + 1101 = (ABCD)_1$ + " "

It is known that this latin square has no orthogonal mate (Hedayat [1969]). This means that no orthogonal partition of the remaining sum of squares can be made which forms a latin square.

If on the other hand, the latin square used is

α	β	γ	δ
β	α	δ	γ
γ	δ	α	β
δ	γ	β	α

, the combinations

corresponding to the Greek letters are:

α: $0000 + 0110 + 1010 + 1111 = (ABCD)_0$ + other effects

β: $0011 + 0101 + 1001 + 1100 = (ABCD)_0$ + " "

γ: $0010 + 0111 + 1000 + 1110 = (ABCD)_1 + (AC)_1$ + other effects

δ: $0001 + 0100 + 1011 + 1101 = (ABCD)_1$ + other effects

This square has two mutually orthogonal mates and hence there must be partitions of the sums of squares into orthogonal components which correspond to the symbols in a latin square.

Instead of inserting symbols in the latin square of order 4, denote the symbols in the latin square by the following partial confounding scheme:

i) add the two 1/8 replicates generated by $((A)_0, (D)_0, (BC)_0)$ and $((A)_1, (C)_1, (ABD)_1)$ to obtain the 4 combinations $(0000 + 0110) + (1010 + 1111)$ and denote these 4 combinations as symbol α,

ii) add the two 1/8 replicates generated by $((D)_1, (AB)_1, (AC)_0)$ and $((AB)_0, (C)_0, (AD)_1)$ to obtain combinations $(0101 + 1011) + (1100 + 0001)$ and denote these 4 combinations as symbol β,

iii) add the two 1/8 replicates generated by $((A)_1, (D)_0, (ABC)_1)$ and $((A)_0, (C)_1, (BD)_0)$ to obtain combinations $(1000 + 1110) + (0010 + 0111)$ and denote these 4 as symbol γ,

iv) add the two 1/8 replicates generated by $((AB)_0, (AC)_1, (D)_1)$ and $((AB)_1, (C)_0, (BD)_1)$ to obtain the combinations $(1101 + 0011) + (0100 + 1001)$ and denote these 4 as symbol δ.

This procedure results in the following latin square of order 4:

α	δ	γ	β
β	α	δ	γ
γ	β	α	δ
δ	γ	β	α

Obviously, one could take any pair of 1/8 replicates such that the 4 combinations are in different rows and in different columns to form the combinations for a given symbol.

The above type of partial confounding results in the class of latin squares denoted as half-plaid latin squares. If partial confounding were utilized in rows as well as in columns the resulting square would be denoted as a plaid latin square (so-called because of its resemblence to plaid cloth if the effects confounded were of different colors). The three types of squares are illustrated below for a latin square of order 6:

Complete confounding of effects

Rows	Columns 1 = $(A)_0,(C)_0$	2 = $(A)_0,(C)_1$	3 = $(A)_0,(C)_2$	4 = $(A)_1,(C)_0$	5 = $(A)_1,(C)_1$	6 = $(A)_2,(C)_2$
1 = $(B)_0,(D)_0$	0000	0010	0020	1000	1010	1020
2 = $(B)_0,(D)_1$	0001	0011	0021	1001	1011	1021
3 = $(B)_0,(D)_2$	0002	0012	0022	1002	1012	1022
4 = $(B)_1,(D)_0$	0100	0110	0120	1100	1110	1120
5 = $(B)_1,(D)_1$	0101	0111	0121	1101	1111	1121
6 = $(B)_1,(D)_2$	0102	0112	0122	1102	1112	1122

Partial confounding of effects with columns

Rows	Columns: 1 = $(C)_0$	2 = $(C)_1$	3 = $(C)_2$	4 = $(CD)_0$	5 = $(CD)_1$	6 = $(CD)_2$
1 = $(B)_0, (D)_0$	0000	0010	0020	1000	1010	1020
2 = $(B)_0, (D)_1$	0001	0011	0021	1021	1001	1011
3 = $(B)_0, (D)_2$	0002	0012	0022	1012	1022	1002
4 = $(B)_1, (D)_0$	1100	1110	1120	0100	0110	0120
5 = $(B)_1, (D)_1$	1101	1111	1121	0121	0101	0111
6 = $(B)_1, (D)_2$	1102	1112	1122	0112	0122	0102

Partial confounding in both rows and columns

Rows	Columns: 1 = $(C)_0$	2 = $(C)_1$	3 = $(C)_2$	4 = $(CD)_0$	5 = $(CD)_1$	6 = $(CD)_2$
1 = $(D)_0$	00	10	20	00	10	20
2 = $(D)_1$	01	11	21	21	01	11
3 = $(D)_2$	02	12	22	12	22	02
4 = $(CD^2)_0$	00	11	22	00	22	11
5 = $(CD^2)_1$	02	10	21	21	10	02
6 = $(CD^2)_2$	01	12	20	12	01	20

In the last table above only the subscripts for combinations of factors c and d have been inserted. There is some difficulty in inserting subscripts for factors a and b such that these effects are orthogonal to both rows and columns. In any event, this problem requires further study to determine if half-plaid latin squares and plaid latin squares lead to latin squares not of the same type as given by complete confounding. If the three types of latin squares of order 6 can be produced by partial and complete confounding, this would be an interesting result.

III. Fractional Replication Construction of O(n,t) Sets

Any latin square may be considered as an n^{-1} fraction of an n^3 factorial where the rows represent levels of one factor, the columns represent the levels of the second factor, and the symbols in the latin square represent the levels of the third factor. As an illustration, consider the latin square of order 3 where the 9 combinations represent the 1/3 fraction of a 3^3 factorial as follows:

Rows	Columns 0	1	2
0	000	012	021
1	102	111	120
2	201	210	222

The above is the 1/3 fraction of a 3^3 corresponding to $(ABC)_{h+i+j=0,\text{ mod } 3}$. Since this is a regular fraction we may write out the aliasing structure in this fraction as follows:

$$\begin{array}{l}
M + ABC \\
A + AB^2C^2 + BC \\
B + AB^2C + AC \\
C + ABC^2 + AB \\
AB^2 + AC^2 + BC^2
\end{array}$$

where the effects connected with a plus sign are completely confounded with each other. In the above latin square the symbols 0,1,2 correspond to the levels of the third factor, c. Now if we set up a second latin square in which the symbols, say α,β,γ, correspond to the levels of AB^2, the resulting square will be orthogonal to the first one. The square corresponding to levels of $(AB^2)_{i+2j,\text{ mod } 3}$ is

$000 + 111 + 222 = \alpha$

$021 + 210 + 102 = \beta$

$201 + 012 + 120 = \gamma$

α	γ	β
β	α	γ
γ	β	α

The class of fractional replicates constituted as an n^{-1} fraction of an n^3 factorial becomes an important one to study as it relates to construction of mutually orthogonal latin squares. In particular, all 2^{-3} fractions of a 2^9 and all 3^{-2} fractions of a 3^6 with all possible aliasing structures could produce several sets of mutually orthogonal latin squares. This could have interesting consequences in finite geometry.

The structure of the left-hand set of parameters in an aliasing structure will have a pattern; for example, for $n = 4$, 5, and 7, the patterns are:

$n = 4$	$n = 5$	$n = 7$
$M + ABC$	$M + ABC$	$M + ABC$
A	A	A
B	B	B
C	C	C
AB^2	AB^2	AB^2
AB^3	AB^3	AB^3
	AB^4	AB^4
		AB^5
		AB^6

Note that although ABC was completely confounded with the mean, any one of the other three-factor interaction components AB^uC^v, $u, v = 1, 2, \ldots, n-1$ could have been utilized equally well. Also, note that the levels of C corresponding to symbols produce a latin square, and that the levels of effects below the factor B produce a set of $n-1$ mutually orthogonal latin squares.

In general we want to look at all possible n^{-1} fractions of an n^3 factorial, i.e., the subset of $\binom{n^3}{n^2}$ combinations for which the levels of C are the symbols in a latin square and to study their patterns especially for $n = 7, 8$, and 9. All possible fractions, or rather all forms of the aliasing structure, could be classified into all types of t mutually orthogonal latin squares, $O(n,t)$ for $t = 1, 2, \ldots, n-1$. Perhaps this is the manner in which the geometries of various values of n can be exhaustively studied. In fractional factorial notation we want to study all possible patterns of $X_{11}^{-1}X_{12}$ in the following matrix equation:

$$\begin{pmatrix} M \\ A \\ B \\ C \\ \vdots \end{pmatrix} + X_{11}^{-1}X_{12}\beta_0 = Y_r$$

where the form of the first vector below the letter C will be determined by the values in $X_{11}^{-1}X_{12}$, the candidates for entry in the vector β_0 are the remaining two- and three-factor interactions, and Y_r is the particular set of n^2 out of n^3 combinations for which the levels of any fourth effect in the first vector form a latin square. Thus, it becomes important to study the properties of $X_{11}^{-1}X_{12}$ even for the 2^m system. The irregular fractions would appear to be the most interesting for $n = 7, 8$, and 9 since regular fractions can be related to complete confounding in section II.1 and to flats and points in the projective geometry.

We now wish to illustrate the use of fractional replication procedures to construct latin squares which are mateless and which have orthogonal mates. To illustrate let us consider the four standard latin squares of order 4 which are (Fisher and Yates [1957]):

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Square I

A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

Square II

A	B	C	D
B	A	D	C
C	D	B	A
D	C	A	B

Square III

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

Square IV

It is known (Hedayat [1969]) that the first three squares are mateless and that the last square belongs to an O(4, 3) set.

Now number the rows as 0, 1, 2, 3 and denote these as levels of the factor a; number the columns as 0, 1, 2, 3, and as 0, 1, 2, 3 for A, B, C, D, respectively, and denote these as levels of the factor c. Then, in factorial notation the above 16 combinations form a one-fourth fraction of a 4^3 factorial treatment design. The aliasing scheme for the fractional replicate given as square IV is

$$
\begin{aligned}
&M + ABC\\
&A + BC + AB^2C^2 + AB^3C^3\\
&B + AC + AB^2C + AB^3C\\
&C + AB + ABC^2 + ABC^3
\end{aligned}
$$

where the effects connected with a plus sign are completely confounded with each other. The completion of the remaining two aliasing structures results in the complete aliasing structures for this 4^{-1} fraction of the 4^3 factorial; these two are:

$$
\begin{aligned}
&AB^2 + AC^3 + BC^2 + AB^3C^2\\
&AB^3 + AC^2 + BC^3 + AB^2C^3
\end{aligned}
$$

If we use the levels of AB^2 and of AB^3 to form two latin squares, these two with square IV form an O(4, 3) set of mutually orthogonal latin squares.

Now, let us return to the set of four standard squares given above and we note that only four combinations in square IV are replaced to obtain squares I, II, and III. These are:

	additional combinations	combinations replaced in IV
Square I	112, 130, 310, 332	110, 132, 312, 330
" II	113, 120, 210, 223	110, 123, 213, 220
" III	213, 230, 320, 331	220, 231, 321, 330

The aliasing structure (without the coefficients) is given on the following page for all four standard latin squares of order 4. The 1/4 replicate given by square IV forms a regular fraction. The remaining three fractional replicates are such that none of the additional effects are unconfounded with the effects A, B, or C of the original latin squares of order 4. Since this is true no linear combinations of these effects will be unconfounded. In order to form a latin square which is orthogonal to the given one it is necessary that there be a set of effects which is unconfounded with the effects in the given square. This is impossible for the three squares I, II, and III and hence the squares are mateless as is well-known.

It would be interesting to ascertain the aliasing structures for the six standard latin squares of order 5 belonging to the O(5,4) set and for the fifty standard latin squares of order 5 for which are known to be mateless (Hedayat [1969]). After a study of these fractions, one should continue such a study for n = 7, 8, and 9. It is suggested that one consider a 2^{6-2} fraction instead of a 4^{3-1} fraction for n = 4 and a 2^{9-3} fraction instead of an 8^{3-1} fraction for n = 8. The reason for this is that there is much more theory available for the

Aliasing structure of effects in the four 1/4 fractional replicates of a 4^3 factorial for four standard latin squares of order 4

	Square I				Square II				Square III				Square IV			
	Effect				Effect				Effect				Effect			
Effect	mean = M	rows = A	cols. = B	letters = C	mean = M	rows = A	cols. = B	letters = C	mean = M	rows = A	cols. = B	letters = C	mean = M	rows = A	cols. = B	letters = C
M	-				-				-				-			
A		-				-				-				-		
B			-				-				-				-	
C				-				-				-				-
AB	P	P	P	P				P				P				C
AB^2				P				P				P				
AB^3				P				P				P				
AC			P				P				P				C	
AC^2			P				P				P					
AC^3			P				P				P					
BC	P	P	P			P				P				C		
BC^2	P	P	P			P				P						
BC^3	P	P	P			P				P						
ABC	P	P	P	P	P	P	P	P		P	P	P	C			
ABC^2	P	P	P	P	P	P	P	P	P	P	P	P				C
ABC^3	P	P	P	P	P	P	P	P	P	P	P	P				C
AB^2C	P	P	P	P	P	P	P	P	P	P	P	P			C	
AB^2C^2	P	P	P	P	P	P	P	P	P	P	P	P		C		
AB^2C^3	P	P	P	P		P	P	P		P	P	P				
AB^3C	P	P	P	P	P	P	P	P	P	P	P	P			C	
AB^3C^2		P	P	P		P	P	P		P	P	P				
AB^3C^3	P	P	P	P	P	P	P	P	P	P	P	P		C		
No. of effects confounded with	12	13	16	12	7	12	12	12	6	12	12	12	1	3	3	

\- means identical effect C means complete confounding

P means partial confounding blank means unconfounded

$s = 2$ in the s^m series than for any other value of s. Also, one may use the generalized defining contrast which has been developed by Raktoe and Federer [1969] to a considerable advantage in writing out aliasing structures in these cases. Investigation of the regular and irregular fractional replicates obtainable for various values of n could lead to considerable advances in the theory of mutually orthogonal latin squares.

IV. ANOVA Construction of O(n,t) Sets

There should be some procedure which would utilize the orthogonality of single degree of freedom contrasts in the analysis of variance (ANOVA) and which could be utilized to construct orthogonal latin squares. For example, one could make use of orthogonal polynomial coefficients for row and column contrasts and then construct mutually orthogonal latin squares from these. To illustrate, consider the latin square of order 4 used previously wherein the row-column intersections are numbered as a 2^4 factorial, i.e.:

	Column			
Row	1	2	3	4
1	0000	0001	0010	0011
2	0100	0101	0110	0111
3	1000	1001	1010	1011
4	1100	1101	1110	1111

The relation between the 16 contrasts using orthogonal polynomial coefficients and the 2^4 factorial is given below:

Source of variation	df		
C.F.M.	1		
Row contrasts	3		
$A = -R_L - 2R_C$	1	1	Rows linear $= R_L = A + 2B$
$B = -2R_L + R_C$	1	1	" quadratic $= R_Q = AB$
$AB = R_Q$	1	1	" cubic $= R_C = 2A - B$
Column contrasts	3		
$C = -C_L - 2C_C$	1	1	Columns linear $= C_L = C + 2D$
$D = -2C_L + C_C$	1	1	" quadratic $= C_Q = CD$
$CD = C_Q$	1	1	" cubic $= C_C = 2C - D$
Roman numbers $= (AB^{u_1})$	3		
$AC = R_LC_L + 4R_CC_C$	1	1	R_LC_L
$BD = 4R_LC_L + R_CC_C$	1	1	R_CC_C
$ABCD = R_QC_Q$	1	1	R_QC_Q
Greek letters $= (AB^{u_2})$	3		
$ABD = -2R_QC_L + R_QC_C$	1	1	R_LC_Q
$BC = 2R_LC_L - 2R_CC_C + 4R_LC_C - R_CC_L$	1	1	R_QC_C
$ACD = (-R_L - 2R_C)C_Q$	1	1	R_CC_L
Latin letters $= (AB^{u_3})$	3		
$AD = 2R_LC_L - 2R_CC_C - R_LC_C + 4R_CC_L$	1	1	R_LC_C
$ABC = R_Q(-C_L - 2C_C)$	1	1	R_QC_L
$BCD = (-2R_L + R_C)C_Q$	1	1	R_CC_Q
Total	16		

The individual degree of freedom contrast matrix for the above 16 combinations is:

Combination

Contrast	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
Mean	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
R_L	−3	−3	−3	−3	−	−	−	−	+	+	+	+	3	3	3	3
R_Q	+	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
R_C	+	+	+	+	−3	−3	−3	−3	3	3	3	3	−	−	−	−
C_L	−3	−	+	3	−3	−	+	3	−3	−	+	3	−3	−	+	3
C_Q	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
C_C	+	−3	3	−	+	−3	3	−	+	−3	3	−	+	−3	3	−
R_LC_L	9	3	−3	−9	3	+	−	−3	−3	−	+	3	−9	−3	3	9
R_LC_Q	−3	3	3	−3	−	+	+	−	+	−	−	+	3	−3	−3	3
R_LC_C	−3	9	−9	3	−	3	−3	+	+	−3	3	−	3	−9	9	−3
R_QC_L	−3	−	+	3	3	+	−	−3	3	+	−	−3	−3	−	+	3
R_QC_Q	+	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+
R_QC_C	+	−3	3	−	−	3	−3	+	−	3	−3	+	+	−3	3	−
R_CC_L	−3	−	+	3	9	3	−3	−9	−9	−3	3	9	3	+	−	−3
R_CC_Q	+	−	−	+	−3	3	3	−3	3	−3	−3	3	−	+	+	−
R_CC_C	+	−3	3	−	−3	9	−9	3	3	−9	9	−3	−	3	−3	+

The corresponding single degree of freedom contrast matrix for the 2^4 factorial is:

Contrast	Combination 0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
Mean	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+
B	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+
AB	+	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
C	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+
D	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+
CD	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
AC	+	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
BD	+	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
ABCD	+	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+
ABD	+	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
BC	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
ACD	−	+	+	−	−	+	+	−	+	−	−	+	+	−	−	+
ABC	−	−	+	+	+	+	−	−	+	+	−	−	−	−	+	+
AD	+	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
BCD	−	+	+	−	+	−	−	+	−	+	+	−	−	+	−	+

The particular contrast matrix utilized is not unique as has been demonstrated above. All orthogonal contrast matrices resulting in latin squares could be considered. For example, other sets of contrasts among rows (or columns) could be:

	1	2	3	4
Mean	+	+	+	+
R_1	-	·	0	0
R_2	0	0	-	+
R_3	+	+	-	-

or

	1	2	3	4
Mean	+	+	+	+
R_1	-	+	0	0
R_2	+	+	-2	0
R_3	+	+	+	-3

The interaction of row and column contrasts possibly could be utilized to allocate the symbols in the latin square.

We wish to illustrate the method of constructing latin squares using orthogonal polynomial coefficients. We shall first consider the construction of three mutually orthogonal latin squares of order 4 and then we shall consider the construction of a single latin square of order 6. In the preceding table on orthogonal polynomials for $n = 4$ denote all combinations with a plus sign as belonging to $(R_L C_L)_1$ and those with a minus sign as belonging to $(R_L C_L)_0$. Do likewise for the $R_Q C_Q$ and $R_C C_C$ effects. Then, the four latin square symbols are obtained as follows:

$$(R_LC_L)_1,\ (R_QC_Q)_1,\ (R_CC_C)_1 = 0000 + 0101 + 1010 + 1111 = A$$

$$(R_LC_L)_1,\ (R_QC_Q)_0,\ (R_CC_C)_0 = 0001 + 0100 + 1011 + 1110 = B$$

$$(R_LC_L)_0,\ (R_QC_Q)_0,\ (R_CC_C)_1 = 0010 + 0111 + 1000 + 1101 = C$$

$$(R_LC_L)_0,\ (R_QC_Q)_1,\ (R_CC_C)_0 = 0011 + 0110 + 1001 + 1100 = D$$

This results in the following latin square of order 4

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

Likewise, if we use the following polynomial contrasts we obtain the two mutually orthogonal mates of the above square:

$$(R_LC_Q)_1,\ (R_QC_C)_0,\ (R_CC_L)_0 = 0001 + 0110 + 1000 + 1111 = \alpha$$

$$(R_LC_Q)_1,\ (R_QC_C)_1,\ (R_CC_L)_1 = 0010 + 0101 + 1011 + 1100 = \beta$$

$$(R_LC_Q)_0,\ (R_QC_C)_0,\ (R_CC_L)_1 = 0011 + 0100 + 1010 + 1101 = \gamma$$

$$(R_LC_Q)_0,\ (R_QC_C)_1,\ (R_CC_L)_0 = 0111 + 1001 + 1110 + 0000 = \delta$$

and

$$(R_LC_C)_1,\ (R_QC_L)_1,\ (R_CC_Q)_1 = 0011 + 0101 + 1000 + 1110 = I$$

$$(R_LC_C)_1,\ (R_QC_L)_0,\ (R_CC_Q)_0 = 0001 + 0111 + 1010 + 1100 = II$$

$$(R_LC_C)_0,\ (R_QC_L)_0,\ (R_CC_Q)_1 = 0000 + 0110 + 1011 + 1101 = III$$

$$(R_LC_C)_0,\ (R_QC_L)_1,\ (R_CC_Q)_0 = 0010 + 0100 + 1001 + 1111 = IV$$

The above results in the following two latin squares of order 4

δ	α	β	γ
γ	β	α	δ
α	δ	γ	β
β	γ	δ	α

III	II	IV	I
IV	I	III	II
I	IV	II	III
II	III	I	IV

The above method of constructing mutually orthogonal latin squares using polynomial coefficients works for latin squares of order n where $n = 2^p$. We need another procedure for other values of n and shall now construct a latin square of order 6 from the orthogonal polynomial coefficients in the table of single degree of freedom contrasts for 36 combinations. If we observe only the signs of contrasts we note that the 36 combinations may be classified into six sets of four with like signs and two additional sets of six. The latter two sets will be used to build up the six sets of four into six sets of six as follows where all combinations with a plus sign go in the one level and all those with a minus sign go in the zero level:

$(R_2C_2)_1$, $(R_3C_3)_1$, $(R_4C_4)_0$, $(R_5C_5)_0$ + 2 from $(R_1C_1)_1, (R_2C_2)_1 (R_3C_3)_1, (R_4C_4)_1, (R_5C_5)_1$

$(R_2C_2)_0$, $(R_3C_3)_0$, $(R_4C_4)_1$, $(R_5C_5)_0$ + "

$(R_2C_2)_0$, $(R_3C_3)_0$, $(R_4C_4)_0$, $(R_5C_5)_0$ + "

$(R_2C_2)_0$, $(R_3C_3)_1$, $(R_4C_4)_0$, $(R_5C_5)_1$ + 2 from $(R_1C_1)_0, (R_2C_2)_1, (R_3C_3)_0, (R_4C_4)_1, (R_5C_5)_0$

$(R_2C_2)_0$, $(R_3C_3)_1$, $(R_4C_4)_1$, $(R_5C_5)_0$ + "

$(R_2C_2)_1$, $(R_3C_3)_0$, $(R_4C_4)_0$, $(R_5C_5)_1$ + "

From these sets we obtain

$$(12 + 21 + 34 + 43) + (00 + 55) = A$$
$$(02 + 20 + 35 + 53) + (11 + 44) = B$$
$$(01 + 10 + 45 + 54) + (22 + 33) = C$$
$$(04 + 15 + 40 + 51) + (23 + 32) = D$$
$$(03 + 25 + 30 + 52) + (14 + 41) = E$$
$$(13 + 24 + 31 + 42) + (05 + 50) = F$$

This results in the following latin square of order 6:

00 A	10 C	20 B	30 E	40 D	50 F
01 C	11 B	21 A	31 F	41 E	51 D
02 B	12 A	22 C	32 D	42 F	52 E
03 E	13 F	23 D	33 C	43 A	53 B
04 D	14 E	24 F	34 A	44 B	54 C
05 F	15 D	25 E	35 B	45 C	55 A

The pair of treatments in the second set of parentheses, e.g. (00 + 55), were picked from the set of six in such a manner as to have i and j in the combination ij, contain 0, 1, 2, 3, 4, and 5 since each letter must appear once in each row and once in each column.

It would be interesting and perhaps enlightening to carry out the above procedure for $n = 10$ and 12 and to exhaustively study the complete set of 35 contrasts for $n = 6$.

	00	01	02	03	04	05	10	11	12	13	14	15
R_1	−5	−5	−5	−5	−5	−5	−3	−3	−3	−3	−3	−3
R_2	5	5	5	5	5	5	−1	−1	−1	−1	−1	−1
R_3	−5	−5	−5	−5	−5	−5	7	7	7	7	7	7
R_4	1	1	1	1	1	1	−3	−3	−3	−3	−3	−3
R_5	−1	−1	−1	−1	−1	−1	5	5	5	5	5	5
C_1	−5	−3	−1	1	3	5	−5	−3	−1	1	3	5
C_2	5	−1	−4	−4	−1	5	5	−1	−4	−4	−1	5
C_3	−5	7	4	−4	−7	5	−5	7	4	−4	−7	5
C_4	1	−3	2	2	−3	1	1	−3	2	2	−3	1
C_5	−1	5	−10	10	−5	1	−1	5	−10	10	−5	1
R_1C_1	25	15	5	−5	−15	−25	15	9	3	−3	−9	−15
R_2C_2	25	−5	−20	−20	−5	25	−5	1	4	4	1	−5
R_3C_3	25	−35	−20	20	35	−25	−35	49	28	−28	−49	35
R_4C_4	1	−3	2	2	−3	1	−3	9	−6	−6	9	−3
R_5C_5	1	−5	10	−10	5	−1	−5	25	−50	50	−25	[illegible]

	20	21	22	23	24	25	30	31	32	33	34	35
R_1	−1	−1	−1	−1	−1	−1	1	1	1	1	1	1
R_2	−4	−4	−4	−4	−4	−4	−4	−4	−4	−4	−4	−4
R_3	4	4	4	4	4	4	−4	−4	−4	−4	−4	−4
R_4	2	2	2	2	2	2	2	2	2	2	2	2
R_5	−10	−10	−10	−10	−10	−10	10	10	10	10	10	10
C_1	−5	−3	−1	1	3	5	−5	−3	−1	1	3	5
C_2	5	−1	−4	−4	−1	5	5	−1	−4	−4	−1	5
C_3	−5	7	4	−4	−7	5	−5	7	4	−4	−7	5
C_4	1	−3	2	2	−3	1	1	−3	2	2	−3	1
C_5	−1	5	−10	10	−5	1	−1	5	−10	10	−5	1
R_1C_1	5	3	1	−1	−3	−5	−5	−3	−1	1	3	5
R_2C_2	−20	4	16	16	4	−20	−20	4	16	16	4	−20
R_3C_3	−20	28	16	−16	−28	20	20	−28	−16	16	28	−20
R_4C_4	2	−6	4	4	−6	2	2	−6	4	4	−6	2
R_5C_5	10	−50	100	−100	50	−10	−10	50	−100	100	−50	10

	40	41	42	43	44	45	50	51	52	53	54	55
R_1	3	3	3	3	3	3	5	5	5	5	5	5
R_2	−1	−1	−1	−1	−1	−1	5	5	5	5	5	5
R_3	−7	−7	−7	−7	−7	−7	5	5	5	5	5	5
R_4	−3	−3	−3	−3	−3	−3	1	1	1	1	1	1
R_5	−5	−5	−5	−5	−5	−5	1	1	1	1	1	1
C_1	−5	−3	−1	1	3	5	−5	−3	−1	1	3	5
C_2	5	−1	−4	−4	−1	5	5	−1	−4	−4	−1	5
C_3	−5	7	4	−4	−7	5	−5	7	4	−4	−7	5
C_4	1	−3	2	2	3	1	1	−3	2	2	−3	1
C_5	−1	5	−10	10	−5	1	−1	5	−10	10	−5	1
R_1C_1	−15	−9	−3	3	9	15	−25	−15	−5	5	15	25
R_2C_2	−5	1	4	4	1	−5	25	−5	−20	−20	−5	25
R_3C_3	35	−49	−28	28	49	−35	−25	35	20	−20	−35	25
R_4C_4	−3	9	−6	−6	9	−3	1	−3	2	2	−3	1
R_5C_5	5	−25	50	−50	25	−5	−1	5	−10	10	−5	1

V. Group Construction of $O(n,t)$ Sets

V.0. Introduction. The construction of $O(n,t)$ sets based on groups and their associated mappings such as automorphism, complete mapping, and orthomorphism is the oldest and still the most popular method for n not of the form $4t+2$. Euler [1782] implicitly utilized some properties of finite groups of order $2t+1$ and $4t$ for his construction of $O(2t+1,2)$ and $O(4t,2)$ sets, respectively. It was MacNeish [1922] who, for the first time, explicitly (however, not rigorously) utilized group properties for his construction of $O(q^m, q^m-1)$ sets and $O(n,\lambda)$ sets, where q is a prime, m is a positive integer and if $n = q_1^{\ell_1} q_2^{\ell_2} \dots q_r^{\ell_r}$ is the prime power decomposition of n then $\lambda = \min(q_1^{\ell_1}, q_2^{\ell_2}, \dots, q_r^{\ell_r})$. The field construction of $O(q^m, q^m-1)$ sets found independently by Bose [1938] and Stevens [1939] is based on the additive group of $GF(q^m)$ and its related cyclic group of automorphisms. The $O(n,n-1)$ sets for $n = 3,4,5,7,8$ and 9 exhibited by Fisher and Yates [1957] are based on cyclic group and abelian groups. Several beautiful applications of group theory to the existence and non-existence of $O(n,t)$ sets have been found by Mann [1942, 1943, 1944]. The $O(12,5)$ sets found by Johnson et al. [1961] and Bose et al. [1960] are based on abelian groups of order 12. Hedayat [1969] and Hedayat and Federer [1969] have found a series of results on the existence and non-existence of $O(n,t)$ sets through the group theory approach. The interested reader on this subject will find the following references together with the references given to these papers very useful: Page [1951], Page-Hall [1955], Singer [1961], Bruck [1951], and Sade [1958].

The author has no doubt that the reader can find many more interesting papers directly or indirectly related this rich subject.

V.1. Definitions and Notations.

There are several forms of definitions of latin squares and orthogonal latin squares. The following forms are useful for the results which will follow:

Definition V.1.1. A latin square of order n on an n-set $\sum$ is an $n \times n$ matrix whose rows and columns are each a permutation of the set $\sum$. Every latin square of order n may therefore be identified with a set of n permutations $(p_1, p_2, \ldots, p_n)$ where p_i is the permutation associated with the ith row.

Definition V.1.2. Let L_i be a latin square of order n on an n-set $\sum_i$, $i = 1, 2, \ldots, t$. Then, the set $S = \{L_1, L_2, \ldots, L_t\}$ is said to be a mutually orthogonal set of t latin squares if the projection of the superimposed form of the t latin squares on any two n-sets $\sum_i$ and $\sum_j$, $i \neq j$, forms a permutation of the cartesian product set of $\sum_i$ and $\sum_j$. Such a set is denoted as an $O(n,t)$ set.

Definition V.1.3. If $L_1 = (P_{11}, P_{12}, \ldots, P_{1n})$ and $L_2 = (P_{21}, P_{22}, \ldots, P_{2n})$ are two latin squares of order n on an n-set $\sum$, then we may define L_1L_2 to be $L_3 = (P_{11}P_{21}, P_{12}P_{22}, \ldots, P_{1n}P_{2n})$. The generalization to the product of $t > 2$ latin squares follows immediately.

V.2.2. Construction of $O(n,t)$ Sets Based on a Group:

We shall divide the problem into three parts based on whether n is a prime, or a mixture of prime powers. The proof of the subsequent results can be found in the references related to this section.

V. 2.1. n = q a prime. Recall that any prime ordered group is cyclic.

Theorem V. 2.1.1. Let $G = \{P_1, P_2, \ldots, P_q\}$ be a cyclic permutation group of degree q and order q. Then, $S_{11} = \{L_1, L_2, \ldots, L_{q-1}\}$ is an $O(q, q-1)$ set, where $L_i = (P_1^i, P_2^i, \ldots, P_q^i)$.

Demonstration V. 2.1.1. Let $q = 5$. Select any arbitrary generator such as $\binom{1\,2\,3\,4\,5}{3\,5\,2\,1\,4}$ which generates a cyclic permutation group G and, hence, a latin square L_1. Then,

L_1 =

3	5	2	1	4
2	4	5	3	1
5	1	4	2	3
4	3	1	5	2
1	2	3	4	5

, L_2 =

2	4	5	3	1
4	3	1	5	2
3	5	2	1	4
5	1	4	2	3
1	2	3	4	5

, L_3 =

5	1	4	2	3
3	5	2	1	4
4	3	1	5	2
2	4	5	3	1
1	2	3	4	5

, L_4 =

4	3	1	5	2
5	1	4	2	3
2	4	5	3	1
3	5	2	1	4
1	2	3	4	5

For those who do not like to work with permutation groups we present the following theorem:

Theorem V. 2.1.2. Let $L(r)$ be an $n \times n$ square with $ri + j \pmod q$ in its (i, j)th cell, $i, j = 0, 1, \ldots, q-1$. Then, $S_{12} = \{L(1), L(2), \ldots, L(q-1)\}$ is an $O(q, q-1)$ set if q is a prime.

Demonstration V. 2.1.2. Let $q = 5$; then,

$L(1)$ =

0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
3	4	0	1	2
4	0	1	2	3

, $L(2)$ =

0	1	2	3	4
2	3	4	0	1
4	0	1	2	3
1	2	3	4	0
3	4	0	1	2

, $L(3)$ =

0	1	2	3	4
3	4	0	1	2
1	2	3	4	0
4	0	1	2	3
2	3	4	0	1

, $L(4)$ =

0	1	2	3	4
4	0	1	2	3
3	4	0	1	2
2	3	4	0	1
1	2	3	4	0

Note that $L(1)$ in theorem V.2.1.2 is based on the cyclic permutation group generated by $\binom{0\ 1\ 2\ \dots\ q-1}{1\ 2\ 3\ \dots\ 0}$ and $L(i) = L^i(1)$, $i = 2,3,\dots,q-1$. Hence theorem V.2.1.2 is a special case of theorem V.2.1.1.

V.2.2. $n = q^m$ <u>where</u> q <u>is a prime and</u> m <u>any positive integer</u>. Note that this case in particular for $m = 1$ includes case 1. We shall present three theorems for this case. The first two are based on cyclic groups and the third one is based on any group which admits an automorphism of order t.

<u>Theorem V.2.2.1.</u> <u>Let</u> $G = \{P_1, P_2, \dots, P_n\}$ <u>be a cyclic permutation group of degree</u> n <u>and order</u> n. <u>Then,</u> $S_{21} = \{L_1, L_2, \dots, L_\lambda\}$ <u>is an</u> $O(n,\lambda)$ <u>set where</u> $n = q^m$ <u>and</u> $\lambda = q-1$.

<u>Demonstration V.2.2.1.</u> Let $n = 3^2 = 9$. Select any arbitrary generator such as $\binom{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{3\ 4\ 5\ 1\ 6\ 7\ 8\ 9\ 2}$ which generates a cyclic permutation group G and hence, a latin square L. Then, since $\lambda = 2$,

$L_1 =$

3	4	5	1	6	7	8	9	2
5	1	6	3	7	8	9	2	4
6	3	7	5	8	9	2	4	1
7	5	8	6	9	2	4	1	3
8	6	9	7	2	4	1	3	5
9	7	2	8	4	1	3	5	6
2	8	4	9	1	3	5	6	7
4	9	1	2	3	5	6	7	8
1	2	3	4	5	6	7	8	9

and $L_2 =$

5	1	6	3	7	8	9	2	4
7	5	8	6	9	2	4	1	3
9	7	2	8	4	1	3	5	6
4	9	1	2	3	5	6	7	8
3	4	5	1	6	7	8	9	2
6	3	7	5	8	9	2	4	1
8	6	9	7	2	4	1	3	5
2	8	4	9	1	3	5	6	7
1	2	3	4	5	6	7	8	9

is an $O(9,2)$ set.

<u>Conjecture</u>. The set S_{21} is orthogonally locked, meaning that there does not exist a latin square L^* such that $S_{21} \cup \{L^*\}$ is an $O(n,\lambda+1)$ set if n is not

a prime. Note that for n even this conjecture is correct since any latin square of even order based on cyclic permutation group is orthogonally mateless.

An analogous theorem to theorem V.2.1.2 for this case is:

Theorem V.2.2.2. Let $L(r)$ be an $n \times n$ square with $ri + j \pmod n$ in its (i,j) cell, $i = 0,1,2,\ldots,n-1$. Then $S_{22} = \{L(1), L(2), \ldots, L(\lambda)\}$ is an $O(n,\lambda)$ set if $n = q^m$ and $\lambda = q-1$.

Demonstration V.2.2.2. Let $n = q = 3^2$ then,

$L(1) =$

0	1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8	0
2	3	4	5	6	7	8	0	1
3	4	5	6	7	8	0	1	2
4	5	6	7	8	0	1	2	3
5	6	7	8	0	1	2	3	4
6	7	8	0	1	2	3	4	5
7	8	0	1	2	3	4	5	6
8	0	1	2	3	4	5	6	7

and $L(2) =$

0	1	2	3	4	5	6	7	8
2	3	4	5	6	7	8	0	1
4	5	6	7	8	0	1	2	3
6	7	8	0	1	2	3	4	5
8	0	1	2	3	4	5	6	7
1	2	3	4	5	6	7	8	0
3	4	5	6	7	8	0	1	2
5	6	7	8	0	1	2	3	4
7	8	0	1	2	3	4	5	6

is an $O(9,2)$ set. Note that theorem V.2.2.2 is a special case of theorem V.2.2.1 viz., $L(1)$ is based on the cyclic permutation group generated by $\begin{pmatrix} 0 & 1 & 2 & \ldots & n-1 \\ 1 & 2 & 3 & \ldots & 0 \end{pmatrix}$ and $L(i) = L^i(1)$, $i = 2,\ldots,\lambda$.

Theorem V.2.2.3. Let $G = \{a_1 = e$ the identity, $a_2, \ldots, a_n\}$ be a group of order n and α an automorphism of order t on G. Then,

1) $S = \{L_1, L_2, \ldots, L_t\}$ is an $O(n,t)$ set, where

$L_i =$

e	a_2	$\cdots$	a_n
$\alpha^i(a_2)$	$\alpha^i(a_2)a_2$	$\cdots$	$\alpha^i(a_2)a_n$
$\alpha^i(a_3)$	$\alpha^i(a_3)a_2$	$\cdots$	$\alpha^i(a_3)a_n$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\alpha^i(a_n)$	$\alpha^i(a_n)a_2$	$\cdots$	$\alpha^i(a_n)a_n$

$i = 1, 2, \ldots, t$.

2) If in particular $t = n-1$, then one can simplify the construction of an $O(n, n-1)$ set from the following key latin square by a cyclic permutation of its last $n-1$ rows.

$L_0 =$

e	$\alpha(x)$	$\alpha^2(x)$	$\cdots$	$\alpha^t(x)$
$\alpha(x)$	$\alpha(x)\alpha(x)$	$\alpha(x)\alpha^2(x)$	$\cdots$	$\alpha(x)\alpha^t(x)$
$\alpha^2(x)$	$\alpha^2(x)\alpha$	$\alpha^2(x)\alpha^2(x)$	$\cdots$	$\alpha^2(x)\alpha^t(x)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\alpha^t(x)$	$\alpha^t(x)\alpha(x)$	$\alpha^t(x)\alpha^2(x)$	$\cdots$	$\alpha^t(x)\alpha^t(x)$

for any x in G except the identity element.

We see, therefore, that by means of theorem V.2.2.3 one can construct an $O(n,t)$ set if we can find a group G and an automorphism α of order t . In particular, if $t = n-1$ the whole task of construction reduces to the construction of L_0 as described above. If $n = q^m$ then because every elementary abelian q-group G of order n admits an automorphism α of order $n-1$, we can construct an $O(q^m, q^m - 1)$ set based on G and α. Here we present a general method of constructing such an automorphism for any $n = q^m$. In particular, we shall exhibit such an automorphism for the following n:

$n = 2^m$, $m = 2, 3, \ldots, 9$

$n = 3^m$, $m = 2, 3, \ldots, 6$

$n = 5^m$, $m = 2, 3, 4$

$n = 7^m$, $m = 2, 3$

$n = 11^2, 13^2, 17^2, 19^2, 23^2, 29^2$, and 31^2.

This will then perhaps be the largest table that has ever been produced so far for $O(n, n-1)$ sets.

Note that there is no loss of generality if we limit ourselves to the following elementary abelian q-group of order $n = q^m$.

$$G^* = \{(b_1 b_2 \ldots b_m),\ b_j = 0, 1, 2, \ldots, q-1,\ j = 1, 2, \ldots, m\}.$$

The binary operation on G^* is addition mod q componentwise, viz., $(b_1 b_2 \ldots b_m) + (b_1' b_2' \ldots b_m') = (c_1 c_2 \ldots c_m)$ where $c_i = b_i + b_i' \pmod q$. Note that the elements of G^* are simply the treatment combinations of m factors each at q levels. The reason why we have chosen this particular elementary abelian q-group is that it has a well-known structure to those who are concerned with experiment design construction. Note also that G^* is the direct product of m Galois fields, each of order q.

The generator set for every elementary abelian q-group of order q^m consists of m elements, and for uniformity, we may choose the following ordered generator set for G^*.

$$g = \{(100\ldots 0), (01,00\ldots 0), \ldots (00\ldots, 010), (00\ldots 01)\}.$$

Note that the structure of every automorphism α on G^* is completely defined if we know the image of each element of g under α. G^* is a vector space of dimension m over $GF(q)$.

Before proceeding further we need the following well-known result:

Theorem V.2.2.4. Let G be an elementary abelian q-group of order $n = q^m$. Then, Auto G is isomorphic to the (multiplicative) group of all non-singular $m \times m$ matrices with entries in the field of integers mod q.

The construction of an automorphism of order $n-1$ for G^* is equivalent to the construction of an $m \times m$ matrix A such that $A^{n-1} = I$ but $A^t \neq I$ if t is not a multiple of $n-1$ over the field of integers mod q.

We know from linear algebra that if ϕ is a linear map on a vector space V and if $x \in V$ such that $x \neq 0$ but $\phi(x) = x$, then 1 is an eigenvalue of ϕ. Moreover, if $\{\lambda_1, \lambda_2, \ldots, \lambda_t\}$ is the set of eigenvalues of ϕ, then $\{\lambda_1^s, \lambda_2^s, \ldots, \lambda_t^s\}$ is the set of eigenvalues of ϕ^s. Therefore, for our problem we must find a linear map on G^* with a set of eigenvalues λ_i having the property that for each i, $\lambda_i^s \neq 1 \pmod q$ for all $s = 1, 2, \ldots, n-2$ and $\lambda_i^{n-1} = 1$. To do so let F be a $GF(q^m)$ and let β be a generator of the multiplicative cyclic group of $GF(q^m)$, i.e. $\beta^i \neq 1$, $i = 1, 2, \ldots, n-2$ while $\beta^{n-1} = 1$. Let $f(x)$ be a monic irreducible polynomial over $GF(q)$ for β. Note that $f(x)$ has degree m. β is sometimes called a primitive root or mark of F. Now, if we let A be the companion matrix for β, then it is easy to see that A has the desired property.

Example

Let us find an automorphism of order 3 for $G^{*} = \{(00), (01), (10), (11)\}$. It is sufficient, by previous arguments, to find a 2×2 matrix A of order 3 over the field of integer mod 2. Let $GF(2^2) = \{0, 1, \beta, \beta + 1\}$ with the following addition $(+)$ and multiplication $(\cdot)$ tables

+	0	1	β	$\beta + 1$
0	0	1	β	$\beta + 1$
1	1	0	$\beta + 1$	β
β	β	$\beta + 1$	0	1
$\beta + 1$	$\beta + 1$	β	1	0

$\cdot$	0	1	β	$\beta + 1$
0	0	0	0	0
1	0	1	β	$\beta + 1$
β	0	β	$\beta + 1$	1
$\beta + 1$	0	$\beta + 1$	1	β

Note that β is a primitive root for $GF(2^2)$ and $f(x) = x^2 + x + 1$ is a monic irreducible polynomial for β, since $f(\beta) \equiv 0 (\mathrm{mod}\ 2)$. The companion matrix associated with $f(x)$ is

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

As a check

$$A^2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \equiv \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \text{ over } GF(2), \quad A^3 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

over $GF(2)$.

Let us now determine the image of the ordered generator set $g = \{(10), (01)\}$ under A.

$$Ag = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} (10) \\ (01) \end{bmatrix} = \begin{bmatrix} 0\,(10) + 1\,(01) \\ 1\,(10) + 1\,(01) \end{bmatrix} = \begin{bmatrix} (01) \\ (11) \end{bmatrix} .$$

Therefore, $A(10) = (01)$, $A(01) = (11)$, and since $(11) = (10) + (01)$, $(00) = 2(10) + 2(01)$, we have $A(11) = (10)$, $A(00) = (00)$.

Now, we have a group G^* of order 4 and an automorphism of order 3 on G^*. We can now construct an $O(4,3)$ set. Since $e = (00)$, and if we let $x = (10)$ in theorem V.2.2.3, we obtain:

$L_0 =$

(00)	$A(10)$	$A^2(10)$	$A^3(10)$
$A(10)$	$A(10)\,A(10)$	$A(10)\,A^2(10)$	$A(10)\,A^3(10)$
$A^2(10)$	$A^2(10)A(10)$	$A^2(10)A^2(10)$	$A^2(10)A^3(10)$
$A^3(10)$	$A^3(10)A(10)$	$A^3(10)A^2(10)$	$A^3(10)A^3(10)$

$=$

(00)	(01)	(11)	(10)
(01)	(00)	(10)	(11)
(11)	(10)	(00)	(01)
(10)	(11)	(01)	(00)

.

The other two latin squares are obtained by a cyclic permutation of the last three rows of L_0. Thus,

$L_1 =$

(00)	(01)	(11)	(10)
(10)	(11)	(01)	(00)
(01)	(00)	(10)	(11)
(11)	(10)	(00)	(01)

(00)	(01)	(11)	(10)
(11)	(10)	(00)	(01)
(10)	(11)	(01)	(00)
(01)	(00)	(10)	(11)

$= L_2$.

To simplify the notation we set $(00) = 1$, $(01) = 2$, $(11) = 3$, $(10) = 4$ to obtain:

$$L_0 = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 2 & 1 & 4 & 3 \\ \hline 3 & 4 & 1 & 2 \\ \hline 4 & 3 & 2 & 1 \\ \hline \end{array}\ , \quad L_1 = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 4 & 3 & 2 & 1 \\ \hline 2 & 1 & 4 & 3 \\ \hline 3 & 4 & 1 & 2 \\ \hline \end{array}\ , \text{ and } L_3 = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 3 & 4 & 1 & 2 \\ \hline 4 & 3 & 2 & 1 \\ \hline 2 & 1 & 4 & 3 \\ \hline \end{array}\ .$$

We are now ready to exhibit a generating matrix of order $n - 1 = q^m - 1$ with entries from $GF(q)$ for those n promised before.

n	Generator	Order	m	Generator	Order
2^2	$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$	3	2^3	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$	7
2^4	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$	15	2^5	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$	31
2^6	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$	63	2^7	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	127
2^8	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$	255	2^9	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$	511

n	Generator	Order	n	Generator	Order
3^2	$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$	8	3^3	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$	26
3^4	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$	80	3^5	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}$	242
3^6	$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$	728	5^3	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix}$	124
5^2	$\begin{bmatrix} 0 & 1 \\ 2 & 2 \end{bmatrix}$	24	7^2	$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$	48
5^4	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 3 & 3 & 0 & 3 \end{bmatrix}$	624	11^2	$\begin{bmatrix} 0 & 1 \\ 3 & 3 \end{bmatrix}$	120
7^3	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix}$	342	13^2	$\begin{bmatrix} 0 & 1 \\ 5 & 5 \end{bmatrix}$	168
17^2	$\begin{bmatrix} 0 & 1 \\ 5 & 5 \end{bmatrix}$	288	19^2	$\begin{bmatrix} 0 & 1 \\ 4 & 4 \end{bmatrix}$	360
23^2	$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$	528	27^2	$\begin{bmatrix} 0 & 1 \\ 3 & 3 \end{bmatrix}$	728
29^2	$\begin{bmatrix} 0 & 1 \\ 2 & 2 \end{bmatrix}$	840	31^2	$\begin{bmatrix} 0 & 1 \\ 2 & 2 \end{bmatrix}$	960

To shed more light on the given procedure we go through another example. Let $n = 2^3$. Then

$$G^* = \{(000), (001), (010), (011), (100), (101), (110), (111)\}$$

$$g = \{(100), (010), (001)\} \quad \text{and} \quad A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} .$$

$$Ag = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} (100) \\ (010) \\ (001) \end{bmatrix} = \begin{bmatrix} (010) \\ (001) \\ (101) \end{bmatrix} .$$

Let x in theorem V.2.2.3 be (100). Then,

$$A(100) = (010) ,$$

$$A^2(100) = (001) ,$$

$$A^3(100) = (101) ,$$

$$A^4(100) = (111) ,$$

$$A^5(100) = (110) ,$$

$$A^6(100) = (011) , \text{ and}$$

$$A^7(100) = (100) .$$

Therefore, we obtain L_0 as follows:

L_0 =

(000)	(010)	(001)	(101)	(111)	(110)	(011)	(100)
(010)	(000)	(011)	(111)	(101)	(100)	(001)	(110)
(001)	(011)	(000)	(100)	(110)	(111)	(010)	(101)
(101)	(111)	(100)	(000)	(010)	(011)	(110)	(001)
(111)	(101)	(110)	(010)	(000)	(001)	(100)	(011)
(110)	(100)	(111)	(011)	(001)	(000)	(101)	(010)
(011)	(001)	(010)	(110)	(100)	(101)	(000)	(111)
(100)	(110)	(101)	(011)	(011)	(010)	(111)	(000)

Setting (000) = 1, (010) = 2, (001) = 3, (101) = 4, (111) = 5, (110) = 6, (011) = 7, (100) = 8, then L_0 in a compact form will be:

$L_0 =$

1	2	3	4	5	6	7	8
2	1	7	5	4	8	3	6
3	7	1	8	6	5	2	4
4	5	8	1	2	7	6	3
5	4	6	2	1	3	8	7
6	8	5	7	3	1	4	2
7	3	2	6	8	4	1	5
8	6	4	3	7	2	5	1

.

Now, we can derive $L_1, L_2, \ldots, L_6$ from L_0 by a cyclic permutation of the last 7 rows of L_0, for example,

$L_1 =$

1	2	3	4	5	6	7	8
8	6	4	3	7	2	5	1
2	1	7	5	4	8	3	6
3	7	1	8	6	5	2	4
4	5	8	1	2	7	6	3
5	4	6	2	1	3	8	7
6	8	5	7	3	1	4	2
7	3	2	6	8	4	1	5

, $L_2 =$

1	2	3	4	5	6	7	8
7	3	2	6	8	4	1	5
8	6	4	3	7	2	5	1
2	1	7	5	4	8	3	6
3	7	1	8	6	5	2	4
4	5	8	1	2	7	6	3
6	8	5	7	3	1	4	2
6	8	5	7	3	1	4	2

,

and so on. Note the way L_1 is derived from L_0: except for the first row of L_0 and L_1, which are identical, the i^{th} row of L_0 becomes the $(i+1)^{th}$ row of L_1, and the last row of L_0 becomes the second row of L_1. In general L_j is derived from L_{j-1} in the same fashion as L_1 is derived from L_0.

V. 2. 3. $n = q_1^{m_1} q_2^{m_2} \ldots q_r^{m_r}$, where q_i is a prime such that $q_i \neq q_j$ if $i \neq j$ and m_i is a positive integer, $i = 1, 2, \ldots, r$.

Theorem V.2.3.1. Let $n = q_1^{m_1}, q_2^{m_2}, \ldots, q_r^{m_r}$ be the prime power decomposition of n. Then, there exists an $O(n, \gamma)$ set based on a group, where $\gamma = \min(q_1^{m_1}, q_2^{m_2}, \ldots, q_r^{m_r}) - 1$.

Construction. Let $n_i = q_i^{m_i}$. Then, by the method of theorem V.2.2.3 construct an $O(n_i, n_i - 1)$ set $S_i = \{L_{i1}, L_{i2}, \ldots, L_{in_i-1}\}$, $i = 1, 2, \ldots, r$. Now, let $S_i^* = \{L_{i1}, L_{i2}, \ldots, L_{i\gamma}\}$, $i = 1, 2, \ldots, r$. Then, $H = \{A_1, A_2, \ldots, A_\gamma\}$ is an $O(n, \gamma)$ set where $A_j = L_{1j} \otimes L_{2j} \otimes \cdots \otimes L_{rj}$. $\otimes$ denotes the Kronecker product operation.

Demonstration V.2.3.1. Let $n = 12 = 2^2 \cdot 3$. Then, $\gamma = 2$,

$$S_1 = L_{11} = \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 2 & 3 & 1 \\ \hline 3 & 1 & 2 \\ \hline \end{array}, \quad L_{12} = \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 3 & 1 & 2 \\ \hline 2 & 3 & 1 \\ \hline \end{array},$$

$$S_2 = L_{21} = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 2 & 1 & 4 & 3 \\ \hline 3 & 4 & 1 & 2 \\ \hline 4 & 3 & 2 & 1 \\ \hline \end{array}, \quad L_{22} = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 4 & 3 & 2 & 1 \\ \hline 2 & 1 & 4 & 3 \\ \hline 3 & 4 & 1 & 2 \\ \hline \end{array}, \quad L_{23} = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 3 & 4 & 1 & 2 \\ \hline 4 & 3 & 2 & 1 \\ \hline 2 & 1 & 4 & 3 \\ \hline \end{array}$$

$S_1^* = \{L_{11}, L_{12}\}$ and S_2^* say $\{L_{21}, L_{22}\}$. Then, the reader can easily verify that

$$H = \{A_1 = L_{11} \otimes L_{21},\ A_2 = L_{12} \otimes L_{22}\}$$

is an $O(12, 2)$ set.

Remark. Let n and γ be the same as in theorem V.2.3.1. Then it can be shown that automorphism method fails to produce more than γ mutually orthogonal latin squares. We shortly show that this inherent defect is due to the mapping not to the group structure.

Definition V, 2, 1. Consider for each positive integer n an abstract group G of order n with binary operation $*$. Let Ω be the collection of all one-to-one mappings of G into itself. Then two maps σ and ψ in Ω are said to be orthogonal if for any g in G,

$$(\sigma Z) * (\psi Z)^{-1} = g$$

has a unique solution Z in G. In particular if σ is an identity map then ψ is said to be an orthomorphism map. A t-subset of Ω is said to be a mutually orthogonal set if every two maps in this t-subset are orthogonal.

Let $L(\cdot)$ be an $n \times n$ square. We make a one-to-one correspondence between the rows of $L(\cdot)$ and the elements of G. Thus, by row x we shall mean the row corresponding to the element x in G. Similarly we make a one-to-one correspondence between the columns of $L(\cdot)$ and the elements of G. The cell of $L(\cdot)$ which occurs in the intersection of row x and column y is called the cell (x,y).

Theorem V, 2, 3, 2. Let σ be in Ω. Put in the cell (x,y) of $L(\cdot)$ the element $(\sigma x) * y$ of G. Call the resulting square $L(\sigma)$. Then $L(\sigma)$ is a latin square of order n on G. Moreover if $\{\sigma_1, \sigma_2, \ldots, \sigma_t\}$ is a set of t mutually orthogonal maps then $\{L(\sigma_1), \ldots, L(\sigma_t)\}$ is an $O(n,t)$ set.

Demonstration V, 2, 3, 2. Let $G = \{0, 1, 2\}$ with the binary operation $x_1 + x_2 = x_3 (\text{mod } 3)$, x_i in G. Then the maps σ and ψ with the following definitions are orthogonal.

$$\sigma(0) = 0 \qquad \psi(0) = 0$$
$$\sigma(1) = 1 \qquad \psi(1) = 2$$
$$\sigma(2) = 2 \qquad \psi(2) = 1 \ .$$

The corresponding latin squares to σ and ψ are:

$$L(\sigma) = \begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline 1 & 2 & 0 \\ \hline 2 & 0 & 1 \\ \hline \end{array} \ , \quad L(\psi) = \begin{array}{|c|c|c|} \hline 0 & 1 & 2 \\ \hline 2 & 0 & 1 \\ \hline 1 & 2 & 0 \\ \hline \end{array}$$

which are orthogonal.

V. 3. Construction of $O(n,t)$ sets based on t different groups of order n

Up to now we have been concerned with the construction of $O(n,t)$ sets using a group of order n which admits certain mappings. In this section we want to show that for some n's and t's one can construct $O(n,t)$ sets based on t different groups each of order n. This approach proved useful because it lead to the construction of an $O(15,3)$ set. We should mention that our motivation to search along these lines has stemmed from the following theorem, with a negative flavor, proved by Mann [1944].

Theorem V.3.1. *It is impossible to construct an $O(5,2)$ set based on two different permutation groups.*

For a while we thought that this theorem might be true for other orders. However, it was found that, fortunately, this is not the case as the following two theorems show:

Theorem V.3.2. *It is possible to construct $O(7,2)$ sets based on two different cyclic permutation groups of order 7.*

Proof: By construction $\{L_1, L_2\}$ is an $O(7,2)$ set where

$$L_1 = \begin{array}{|c|c|c|c|c|c|c|}\hline 1&2&3&4&5&6&7\\\hline 3&7&6&1&4&2&5\\\hline 6&5&2&3&1&7&4\\\hline 2&4&7&6&3&5&1\\\hline 7&1&5&2&6&4&3\\\hline 5&3&4&7&2&1&6\\\hline 4&6&1&5&7&3&2\\\hline\end{array}\;, \qquad L_2 = \begin{array}{|c|c|c|c|c|c|c|}\hline 1&2&3&4&5&6&7\\\hline 2&3&4&5&6&7&1\\\hline 3&4&5&6&7&1&2\\\hline 4&5&6&7&1&2&3\\\hline 5&6&7&1&2&3&4\\\hline 6&7&1&2&3&4&5\\\hline 7&1&2&3&4&5&6\\\hline\end{array}\;.$$

L_1 and L_2 are based on two different permutation groups as can easily be seen from the different structure of their rows. To be specific L_1 is based on the cyclic permutation group generated by $\binom{1\ 2\ 3\ 4\ 5\ 6\ 7}{3\ 7\ 6\ 1\ 4\ 2\ 5}$ and L_2 is based on the cyclic permutation group generated by $\binom{1\ 2\ 3\ 4\ 5\ 6\ 7}{2\ 3\ 4\ 5\ 6\ 7\ 1}$. Note that, since L_1 and L_2 are based on cyclic permutation groups, then by theorem V.2.1.1 $\{L_1\}$ and $\{L_2\}$ can be embedded in $O(7,6)$ sets. However, whether or not $\{L_1, L_2\}$ can be embedded in a larger set is an open problem.

Theorem V.3.3. It is possible to construct $O(15, 3)$ sets based on three different cyclic permutation groups of order 15.

We remind the reader that every group of order 15 is cyclic.

Proof: By construction $\{L_1, L_2, L_3\}$ is an $O(15, 3)$ set where

$L_1 =$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	3	4	5	6	7	8	9	10	11	12	13	14	0	1
4	5	6	7	8	9	10	11	12	13	14	0	1	2	3
6	7	8	9	10	11	12	13	14	0	1	2	3	4	5
8	9	10	11	12	13	14	0	1	2	3	4	5	6	7
10	11	12	13	14	0	1	2	3	4	5	6	7	8	9
12	13	14	0	1	2	3	4	5	6	7	8	9	10	11
14	0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2	3	4	5	6	7	8	9	10	11	12	13	14	0
3	4	5	6	7	8	9	10	11	12	13	14	0	1	2
5	6	7	8	9	10	11	12	13	14	0	1	2	3	4
7	8	9	10	11	12	13	14	0	1	2	3	4	5	6
9	10	11	12	13	14	0	1	2	3	4	5	6	7	8
11	12	13	14	0	1	2	3	4	5	6	7	8	9	10
13	14	0	1	2	3	4	5	6	7	8	9	10	11	12

generated by $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 0 & 1 \end{pmatrix}$,

$L_2 =$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	11	10	7	9	14	13	6	0	3	4	2	8	5	12
11	2	4	6	3	12	5	13	1	7	9	10	0	14	8
2	10	9	13	7	8	14	5	11	6	3	4	1	12	0
10	4	3	5	6	0	12	14	2	13	7	9	11	8	1
4	9	7	14	13	1	8	12	10	5	6	3	2	0	11
9	3	6	12	5	11	0	8	4	14	13	7	10	1	2
3	7	13	8	14	2	1	0	9	12	5	6	4	11	10
7	6	5	0	12	10	11	1	3	8	14	13	9	2	4
6	13	14	1	8	4	2	11	7	0	12	5	3	10	9
13	5	12	11	0	9	10	2	6	1	8	14	7	4	3
5	14	8	2	1	3	4	10	13	11	0	12	6	9	7
14	12	0	10	11	7	9	4	5	2	1	8	13	3	6
12	8	1	4	2	6	3	9	14	10	11	0	5	7	13
8	0	11	9	10	13	7	3	12	4	2	1	14	6	5

generated by $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ 1 & 11 & 10 & 7 & 9 & 14 & 13 & 6 & 0 & 3 & 4 & 2 & 8 & 5 & 12 \end{pmatrix}$, and

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
7	4	8	9	11	2	5	12	13	10	0	14	1	3	6
12	11	13	10	14	8	2	1	3	0	7	6	4	9	5
1	14	3	0	6	13	8	4	9	7	12	5	11	10	2
4	6	9	7	5	3	13	11	10	12	1	2	14	0	8
11	5	10	12	2	9	3	14	0	1	4	8	6	7	13
14	2	0	1	8	10	9	6	7	4	11	13	5	12	3
6	8	7	4	13	0	10	5	12	11	14	3	2	1	9
5	13	12	11	3	7	0	2	1	14	6	9	8	4	10
2	3	1	14	9	12	7	8	4	6	5	10	13	11	0
8	9	4	6	10	1	12	13	11	5	2	0	3	14	7
13	10	11	5	0	4	1	3	14	2	8	7	9	6	12
3	0	14	2	7	11	4	9	6	8	13	12	10	5	1
9	7	6	8	12	14	11	10	5	13	3	1	0	2	4
10	12	5	13	1	6	14	0	2	3	9	4	7	8	11

$L_3 =$ (the table above)

generated by $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\ 7 & 4 & 8 & 9 & 11 & 2 & 5 & 12 & 13 & 10 & 0 & 14 & 1 & 3 & 6 \end{pmatrix}$.

Whether or not $\{L_1, L_2, L_3\}$ can be embedded in an $O(15, t)$, $t > 3$, set is an open problem.

V. 4. Concluding Remark

Johnson et al. [1961] and Bose et al. [1960] independently found, by an electronic computer, five mutually orthogonal latin squares by first finding five mutually orthogonal maps for an abelian group of order 12. The $O(12, 5)$ set exhibited below is the set found by Johnson et al. [1961]. Note that the top square is obtained, after a proper renaming, as the direct product of a latin square of order 2 and a cyclic latin square of order 6 being both orthogonally mateless. Moreover, every other square is obtained by proper row permutations, determined by an orthomorphism, from the top square.

Final Remark. The group method fails to produce an $O(n,t)$ set, $t \geq 2$ for any n of the form $4t + 2$. This is so because the Cayley table of any group of order $n = 4t + 2$, which is a latin square of order n, is orthogonally mateless.

0	1	2	3	4	5	6	7	8	9	10	11
5	0	1	2	3	4	11	6	7	8	9	10
4	5	0	1	2	3	10	11	6	7	8	9
3	4	5	0	1	2	9	10	11	6	7	8
2	3	4	5	0	1	8	9	10	11	6	7
1	2	3	4	5	0	7	8	9	10	11	6
6	7	8	9	10	11	0	1	2	3	4	5
11	6	7	8	9	10	5	0	1	2	3	4
10	11	6	7	8	9	4	5	0	1	2	3
9	10	11	6	7	8	3	4	5	0	1	2
8	9	10	11	6	7	2	3	4	5	0	1
7	8	9	10	11	6	1	2	3	4	5	0

0	1	2	3	4	5	6	7	8	9	10	11
6	7	8	9	10	11	0	1	2	3	4	5
10	11	6	7	8	9	4	5	0	1	2	3
4	5	0	1	2	3	10	11	6	7	8	9
11	6	7	8	9	10	5	0	1	2	3	4
5	0	1	2	3	4	11	6	7	8	9	10
9	10	11	6	7	8	3	4	5	0	1	2
7	8	9	10	11	6	1	2	3	4	5	0
2	3	4	5	0	1	8	9	10	11	6	7
8	9	10	11	6	7	2	3	4	5	0	1
1	2	3	4	5	0	7	8	9	10	11	6
3	4	5	0	1	2	9	10	11	6	7	8

0	1	2	3	4	5	6	7	8	9	10	11
3	4	5	0	1	2	9	10	11	6	7	8
6	7	8	9	10	11	0	1	2	3	4	5
5	0	1	2	3	4	11	6	7	8	9	10
9	10	11	6	7	8	3	4	5	0	1	2
7	8	9	10	11	6	1	2	3	4	5	0
4	5	0	1	2	3	10	11	6	7	8	9
10	11	6	7	8	9	4	5	0	1	2	3
1	2	3	4	5	0	7	8	9	10	11	6
2	3	4	5	0	1	8	9	10	11	6	7
11	6	7	8	9	10	5	0	1	2	3	4
8	9	10	11	6	7	2	3	4	5	0	1

0	1	2	3	4	5	6	7	8	9	10	11
10	11	6	7	8	9	4	5	0	1	2	3
5	0	1	2	3	4	11	6	7	8	9	10
7	8	9	10	11	6	1	2	3	4	5	0
1	2	3	4	5	0	7	8	9	10	11	6
9	10	11	6	7	8	3	4	5	0	1	2
3	4	5	0	1	2	9	10	11	6	7	8
8	9	10	11	6	7	2	3	4	5	0	1
4	5	0	1	2	3	10	11	6	7	8	9
11	6	7	8	9	10	5	0	1	2	3	4
6	7	8	9	10	11	0	1	2	3	4	5
2	3	4	5	0	1	8	9	10	11	6	7

0	1	2	3	4	5	6	7	8	9	10	11
2	3	4	5	0	1	8	9	10	11	6	7
7	8	9	10	11	6	1	2	3	4	5	0
8	9	10	11	6	7	2	3	4	5	0	1
4	5	0	1	2	3	10	11	6	7	8	9
11	6	7	8	9	10	5	0	1	2	3	4
10	11	6	7	8	9	4	5	0	1	2	3
6	7	8	9	10	11	0	1	2	3	4	5
9	10	11	6	7	8	3	4	5	0	1	2
5	0	1	2	3	4	11	6	7	8	9	10
3	4	5	0	1	2	9	10	11	6	7	8
1	2	3	4	5	0	7	8	9	10	11	6

VI. Projecting Diagonals Construction of $O(n,t)$ Sets

A very simple procedure (sort of the "man-on-the-street" approach) of constructing balanced incomplete block and partially balanced incomplete block designs for $v = k^2$ items in incomplete blocks of size k has been utilized since the late 1940's by the author and has its counterpart in constructing $O(n,t)$ sets. First we shall illustrate its use in incomplete block experiment design construction, and then we show how it applies to the construction on $O(n,t)$ set. The theoretical basis for this method may be derived directly from the preceding section.

The procedure becomes apparent through an example. Suppose that $v = 9$ and $k = 3$. After writing the first square as illustrated below, take successive diagonals of the preceding square and use them to form the incomplete blocks of a square, thus:

Square 1

1	2	3
4	5	6
7	8	9

Square 2

1	5	9
2	6	7
3	4	8

Square 3

1	6	8
2	4	9
3	5	7

Square 4

1	4	7
2	5	8
3	6	9

As we have noted this is a resolvable balanced incomplete design with the parameters $v = 9 = k^2$, $k = 3$, $r = 4 = k + 1$, $b = 12 = k(k+1)$, and $\lambda = 1$, where the rows of the above squares form the incomplete blocks.

To form a partially balanced incomplete block design for $v = k^2$ in incomplete blocks of size k one may use any 2, any 3, ..., any k arrangements (or squares). To illustrate the formation of a partially balanced incomplete block design for $v = 6 = k(k-1)$, $r = 2$ or $3 = k$, and $k' = 2$ simply delete the numbers 7, 8, and 9 from the last $k = 3$ arrangements. The deletion of certain symbols from the set $1, 2, \ldots, v$ is known as "variety cutting". For $k^2 = 25$ and $k = 5$ partially balanced incomplete block designs may be constructed for $v = 10$ and $k^* = 2$, $v = 15$ and $k^+ = 3$, and $v = 20$ and $k' = 4$ by the above "variety cutting" procedure.

Also, the successive diagonals method is useful for $v = k^2$ in incomplete blocks of size k for any odd k. For example, for $v = 225$ and $k = 15$ four arrangements or squares may be quickly constructed by the above method. Likewise, the "variety cutting" procedure may be utilized to obtain 2 or 3 arrangements for $v = 15p$, $2 \leq p \leq 15$, varieties.

The above method has its counterpart in constructing mutually orthogonal latin squares and this possibility is briefly mentioned in Fisher and Yates [1957] in this context. Again the method becomes apparent through an example. First write the latin square in standard order and of the form given below for the first square, then project the main right diagonal of the preceding square into the first column of a square, and then write the symbol order in the same manner as in the first square. As a first example, let the order of the latin square be 3; the squares are:

first square

1	2	3
2	3	1
3	1	2

second square

1	2	3
3	1	2
2	3	1

third square

1	2	3
1	2	3
1	2	3

Thus, the main right diagonal of the first square is 1,3,2 which becomes the first column of the second square. Then, write the first row as 1,2,3, the second row as 3,1,2, and the third row as 2,3,1. For the third square, which is not a latin square, the right main diagonal of the second square is 1,1,1 and this becomes the first column of the third square; the rows are then completed. If we then take the right main diagonal of the third square, we obtain the first square.

As a second illustrative example, the five squares for order $n = 5$ which are constructed by successively projecting diagonals, are:

first square

1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

second square

1	2	3	4	5
3	4	5	1	2
5	1	2	3	4
2	3	4	5	1
4	5	1	2	3

third square

1	2	3	4	5
4	5	1	2	3
2	3	4	5	1
5	1	2	3	4
3	4	5	1	2

fourth square

1	2	3	4	5
5	1	2	3	4
4	5	1	2	3
3	4	5	1	2
2	3	4	5	1

fifth square

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

The fifth square is not a latin square but may be utilized to construct the first square through use of the method of successive projections of the main diagonals.

The method may be utilized for any odd order n and will produce $q_1 - 1$ orthogonal latin squares for $n = q_1 \cdot q_2 \cdot \ldots \cdot q_s$ where $q_i < q_{i+1}$ and $q_1, q_2, \ldots, q_s$

is the prime power decomposition of n. Thus, for $n = 15 = 3(5)$ a pair $(q_1 - 1 = 3 - 1 = 2)$ of orthogonal latin squares is easily produced. For $n = 35 = 5(7)$, a quartet of mutually orthogonal latin squares is readily produced by the projecting diagonals method.

VII. Relating Between Complete Confounding and Simple Orthomorphisms

We shall illustrate the ideas by going through a complete example taking $n = 12 = 2^2 \times 3$. For this purpose we take the ring of 12 elements (obtained by utilizing Raktoe's [1969] results) as follows:

$GF(2^2)$		I_3	$GF(3)$		I_4
0		0	0		0
1	$\Rightarrow$	3	1	$\Rightarrow$	4
x		3x	2		2
x+1		3x+3			

$R_{12} = I_3 \oplus I_4 = \{0, 1, 2, 3, 4, 5, 3x, 3x+1, 3x+2, 3x+3, 3x+4, 3x+5\}$

R_{12} is a commutative ring under addition and multiplication (modddd6, $3x^2 + 3x + 3$, $4x + 4$) in the following sense:

e.g.: (a). $(3x+3) + (3x+4) = 6x + 7 = 1$; here we have to reduce only mod 6 to get the answer.

(b). $(3x+1) \cdot (3x+4) = 9x^2 + 15x + 4$

$= 3x^2 + 3x + 4$

$= [3+0]\,x^2 + [3+0]\,x + [0+4] = (3x^2 + 3x) + 4$

$= 3 + 1 = 1$; here first we had to reduce mod 6, then mod $3x^2 + 3x + 3$ leaving us immediately 3 and 4, which is irreducible mod $4x + 4$, thus resulting in 1.

Explicitly, to facilitate arithmetic, the addition and mutliplication of these 12 elements are:

+	0	1	2	3	4	5	3x	3x+1	3x+2	3x+3	3x+4	3x+5
0	0	1	2	3	4	5	3x	3x+1	3x+2	3x+3	3x+4	3x+5
1		2	3	4	5	0	3x+1	3x+2	3x+3	3x+4	3x+5	3x
2			4	5	0	1	3x+2	3x+3	3x+4	3x+5	3x	3x+1
3				0	1	2	3x+3	3x+4	3x+5	3x	3x+1	3x+2
4					2	3	3x+4	3x+5	3x	3x+1	3x+2	3x+3
5						4	3x+5	3x	3x+1	3x+2	3x+3	3x+4
3x							0	1	2	3	4	5
3x+1								2	3	4	5	0
3x+2									4	5	0	1
3x+3										0	1	2
3x+4											2	3
3x+5												4

·	0	1	2	3	4	5	3x	3x+1	3x+2	3x+3	3x+4	3x+5
0	0	0	0	0	0	0	0	0	0	0	0	0
1		1	2	3	4	5	3x	3x+1	3x+2	3x+3	3x+4	3x+5
2			4	0	2	4	0	2	4	0	2	4
3				3	0	3	3x	3x+3	3x	3x+3	3x	3x+3
4					4	2	0	4	2	0	4	2
5						1	3x	3x+5	3x+4	3x+3	3x+2	3x+1
3x							3x+3	3	3x+3	3	3x+3	3
3x+1								3x+4	5	3x	1	3x+2
3x+2									3x+1	3	3x+5	1
3x+3										3x+3	3	3x
3x+4											3x+1	5
3x+5												3x+4

Now, associate with a latin square of order 12 the $3^2 \times 4^2 = [3 \times 4] \times [3 \times 4] = 12 \times 12$ lattice square with the following breakdown of the 143 degrees of freedom:

A^4	2	C^3	3
B^4	2	D^3	3
A^4B^4	2	C^3D^3	3
A^4B^2	2	C^3D^{3x}	3
		C^3D^{3x+3}	3

A^4C^3	6	B^4C^3	6	$A^4B^4C^3$	6	$A^4B^2C^3$	6
A^4D^3	6	B^4D^3	6	$A^4B^4D^3$	6	$A^4B^2D^3$	6
$A^4C^3D^3$	6	$B^4C^3D^3$	6	$A^4B^4C^3D^3$	6	$A^4B^2C^3D^3$	6
$A^4C^3D^{3x}$	6	$B^4C^3D^{3x}$	6	$A^4B^4C^3D^{3x}$	6	$A^4B^2C^3D^{3x}$	6
$A^4C^3D^{3x+3}$	6	$B^4C^3D^{3x+3}$	6	$A^4B^4C^3D^{3x+3}$	6	$A^4B^2C^3D^{3x+3}$	6

For any row or column confounding we need to confound effects totaling up to 11 degrees of freedom. There are natural candidates available. In fact, we may choose for our first lattice square the confounding scheme in many ways. A scheme resulting in a pair of orthogonal latin squares is the following:

Confounding scheme of our 12 × 12 lattice square

	$(A^4C^3)_0$	$(A^4C^3)_1$	$(A^4C^3)_2$	$(A^4C^3)_3$	$(A^4C^3)_4$	$(A^4C^3)_5$	$(A^4C^3)_{3x}$	$(A^4C^3)_{3x+1}$	$(A^4C^3)_{3x+2}$	$(A^4C^3)_{3x+3}$	$(A^4C^3)_{3x+4}$	$(A^4C^3)_{3x+5}$
$(B^4D^3)_0$	0000	4030	2000	0030	4000	2030	00 3x 0	40 3x+3 0	20 3x 0	00 3x+3 0	40 3x 0	20 3x+3 0
$(B^4D^3)_1$	0403	4433	2403	0433	4403	2433	04 3x 3	44 3x+3 3	24 3x 3	04 3x+3 3	44 3x 3	24 3x+3 3
$(B^4D^3)_2$	0200	4230	2200	0230	4200	2230	02 3x 0	42 3x+3 0	22 3x 0	02 3x+3 0	42 3x 0	22 3x+3 0
$(B^4D^3)_3$	0003	4033	2003	0033	4003	2033	00 3x 3	40 3x+3 3	20 3x 3	00 3x+3 3	40 3x 3	20 3x+3 3
$(B^4D^3)_4$	0400	4430	2400	0430	4400	2430	04 3x 0	44 3x+3 0	24 3x 0	04 3x+3 0	44 3x 0	24 3x+3 0
$(B^4D^3)_5$	0203	4233	2203	0233	4203	2233	02 3x 3	42 3x+3 3	22 3x 3	02 3x+3 3	42 3x 3	22 3x+3 3
$(B^4D^3)_{3x}$	000 3x	403 3x	200 3x	003 3x	400 3x	203 3x	00 3x 3x	40 3x+3 3x	20 3x 3x	00 3x+3 3x	40 3x 3x	20 3x+3 3x
$(B^4D^3)_{3x+1}$	040 3x+3	443 3x+3	240 3x+3	043 3x+3	440 3x+3	243 3x+3	04 3x 3x+3	44 3x+3 3x+3	24 3x 3x+3	04 3x+3 3x+3	44 3x 3x+3	24 3x+3 3x+3
$(B^4D^3)_{3x+2}$	020 3x	423 3x	220 3x	023 3x	420 3x	223 3x	02 3x 3x	42 3x+3 3x	22 3x 3x	02 3x+3 3x	42 3x 3x	22 3x+3 3x
$(B^4D^3)_{3x+3}$	000 3x+3	403 3x+3	200 3x+3	003 3x+3	400 3x+3	203 3x+3	00 3x 3x+3	40 3x+3 3x+3	20 3x 3x+3	00 3x+3 3x+3	40 3x 3x-3	20 3x+3 3x+3
$(B^4D^3)_{3x+4}$	040 3x	443 3x	240 3x	043 3x	440 3x	243 3x	04 3x 3x	44 3x+3 3x	24 3x 3x	04 3x+3 3x	44 3x 3x	24 3x+3 3x
$(B^4D^3)_{3x+5}$	020 3x+3	423 3x+3	220 3x+3	023 3x+3	420 3x+3	223 3x+3	02 3x 3x+3	42 3x+3 3x+3	22 3x 3x+3	02 3x+3 3x+3	42 3x 3x+3	22 3x+3 3x+3

LATIN SQUARE 1: Treatments identified with $A^4B^4C^3D^3$

0	1	2	3	4	5	3x	3x+1	3x+2	3x+3	3x+4	3x+5
1	2	3	4	5	0	3x+1	3x+2	3x+3	3x+4	3x+5	3x
2	3	4	5	0	1	3x+2	3x+3	3x+4	3x+5	3x	3x+1
3	4	5	0	1	2	3x+3	3x+4	3x+5	3x	3x+1	3x+2
4	5	0	1	2	3	3x+4	3x+5	3x	3x+1	3x+2	3x+3
5	0	1	2	3	4	3x+5	3x	3x+1	3x+2	3x+3	3x+4
3x	3x+1	3x+2	3x+3	3x+4	3x+5	0	1	2	3	4	5
3x+1	3x+2	3x+3	3x+4	3x+5	3x	1	2	3	4	5	0
3x+2	3x+3	3x+4	3x+5	3x	3x+1	2	3	4	5	0	1
3x+3	3x+4	3x+5	3x	3x+1	3x+2	3	4	5	0	1	2
3x+4	3x+5	3x	3x+1	3x+2	3x+3	4	5	0	1	2	3
3x+5	3x	3x+1	3x+2	3x+3	3x+4	5	0	1	2	3	4

LATIN SQUARE 2: Treatments identified with $A^4B^2C^3D^{3x}$

0	1	2	3	4	5	3x	3x+1	3x+2	3x+3	3x+4	3x+5
3x+2	3x+3	3x+4	3x+5	3x	3x+1	2	3	4	5	0	1
4	5	0	1	2	3	3x+4	3x+5	3x	3x+1	3x+2	3x+3
3x	3x-1	3x+2	3x+3	3x+4	3x+5	0	1	2	3	4	5
2	3	4	5	0	1	3x+2	3x+3	3x+4	3x+5	3x	3x+1
3x+4	3x+5	3x	3x-1	3x-2	3x+3	4	5	0	1	2	3
3x+3	3x+4	3x+5	3x	3x+1	3x-2	3	4	5	0	1	2
5	0	1	2	3	4	3x+5	3x	3x+1	3x+2	3x+3	3x+4
3x+1	3x+2	3x+3	3x+4	3x+5	3x	1	2	3	4	5	0
3	4	5	0	1	2	3x+3	3x+4	3x+5	3x	3x+1	3x+2
3x+5	3x	3x+1	3x+2	3x+3	3x+4	5	0	1	2	3	4
1	2	3	4	5	0	3x+1	3x+2	3x+3	3x+4	3x+5	3x

Using the complete confounding approach as outlined above, one can construct $\min[(2^2-1), (3-1)] = 2$ mutually orthogonal latin squares and no more as can easily be observed from the degrees of freedom table.

From the multiplication table of our ring R_{12}, we observe that 1, 5, 3x+1, 3x+2, 3x+4 and 3x+5 are the 6 non-zero divisors (i.e. elements with multiplicative inverses). Following Bose, Chakravarti and Knuth [1960], we consider simple automorphisms in R_{12} of the form:

$$\alpha(r) = r^* r$$

where r^* is a given fixed element having a multiplicative inverse (because only these elements are capable of producing automorphisms of R_{12}). Let now our aim be to produce two orthomorphisms which in turn will produce an $O(12, 2)$ set. For this purpose consider the automorphisms:

$$I(r) = r$$

$$\alpha(r) = r^* \cdot r \ .$$

Now $I-\alpha$ implies the condition that r in the equation $[r^* \cdot r - r] = c$ has a unique solution for every c of R_{12}. In our setting this means that $r(r^* - 1) = c$ has a unique solution, i.e., $r(r^* + 5) = c$ has a unique solution which in turn implies that $(r^* + 5)^{-1}$ exists in R_{12}.

Substituting in the values of r^* we see that:

$[1 + 5]^{-1}$ does not exist in R_{12}

$[5 + 5]^{-1}$ " " " " "

$[(3x+1) + 5]^{-1}$ does not exist in R_{12}

$$[(3x+2)+5]^{-1} = [3x+1]^{-1} \text{ exists in } R_{12}$$

$$[(3x+4)+5]^{-1} \text{ does not exist in } R_{12}$$

$$[(3x+5)+5]^{-1} = [3x+4]^{-1} \text{ exists in } R_{12} .$$

Hence we have obtained two pairs of orthomorphisms namely:

$$I(r) = r \qquad\qquad I(r) = r$$

$$\alpha_1(r) = (3x+2)r \quad \text{and} \quad \alpha_2(r) = (3x+5)r .$$

The $O(12, 2)$ set presented above using complete confounding corresponds to the first pair of maps. It may be easily shown that simple maps of the type $\alpha(r) = r^* \cdot r$ lead to $O(12, 2)$ sets or in general to an $O(n, a)$ set, where $a = \min(p_1^{n_1}-1, p_2^{n_2}-1, \ldots, p_k^{n_k}-1)$ and $n = \prod_{i=1}^{k} p_i^{n_i}$ so that the complete confounding approach is equivalent to the construction of a set of a orthomorphisms.

VIII. Some Remarks on "Orthomorphism" Construction of $O(n,t)$ Sets

In [illegible] and in 1960 E. T. Parker showed by a combination of classification of cases (with considerable elimination of isomorphic repetition accessible to cut down on computer time, followed by computer runs) that there are obtainable only 5 orthogonal latin squares of order 12, all restricted to be copies of the non-cyclic Abelian group with latin squares related by row permutations. (The researchers cited call this the method of orthomorphisms. Parker considers this no method, but only based on freaks of luck; further, Parker feels that "orthomorphism" admits no precise definition.)

Parker made another finding, also by hard classification of cases followed by computer runs, which Marshall Hall feels is more important than that cited above. No pair of order-12 orthogonal squares of the type mentioned can be extended to a complete set of any sort; i.e., further orthogonal latin squares are allowed to be completely general.

What might be obtainable for orthogonal squares of order 20 in like fashion, row-permuting the non-cyclic Abelian group of order 20 is an interesting matter for speculation — conceivably one might even produce a complete set (19 orthogonal squares equivalent to a plane). Knuth and Parker discussed the problem about 19[illegible]3, and concluded that exhaustive search is out of the question; still a fortunate sample of cases might produce an attractive result.

In 196[illegible] Parker looked at the row-permutation ("orthomorphism" of Bose and [illegible]) approach for the group of order 15, and proved by Hasse-Minkowski invariants that a complete set could not be so obtained. He dropped further work; but some persistence could quite possibly yield such as five orthogonal squares of order 15 .

A hybrid attack on order 15 or 20 might be undertaken by an ambitious investigator. (The facts for order 12 mentioned above rule out chances here.) One might produce sets of orthogonal latin squares of row-permuted group type, using automorphisms of the group latin squares to eliminate — or, that failing, reduce — isomorphic repetition. It would not be shrewd to program a computer to produce all transversals of a group latin square, for running time and output would be excessive; then for any hint of efficiency it would be necessary to turn about and do a reduction on the computer output. After a set of row-permuted latin squares (possibly exhaustive for order 15, but almost certainly only a sample for order 20) large enough that computer searching would require realistic amounts of time, one might proceed with the next step. Produce all transversals of the set of orthogonal squares by computer, then fit these together in all possible ways (again by computer) to form orthogonal mates of the produced set of orthogonal latin squares. Unlike Parker's assertion above about complete sets of order-15 squares, there is no known argument implying impossibility of producing 14 orthogonal latin squares of order 15 by this hybrid attack.

It is well-known that removing one line from the plane, usually called the line of infinity, the remaining $n^2 + n$ lines can be arranged into $2n$ lines passing through two points at infinity which are arbitrary up to notation and coordinatization of the plane, and $n^2 - n$ lines belonging to $n-1$ mutually orthogonal latin squares. If the line at infinity is chosen to be a secant and there are $2n$ lines, the lines pass through the two points of the oval such that each of the $n-1$ latin squares consists of $\frac{n}{2}$ secants and $\frac{n}{2}$ non-intersectors passing through each of the $n-1$ points at infinity other than the points of the oval.

Using the described method, it was assumed that a plane of order 10 exists. Under this assumtion 21 lines could be exhibited arbitrarily up to notation. Out of these lines one was taken to be the line of infinity and the remaining 20 used to coordinatize the plane. Then by trial and error twenty more lines were found which formed two orthogonal latin squares. The method used to construct these squares differs from the one described in literature.

Unfortunately no more squares could be found using this [illegible]

computer established that the two squares did not yield an additional mutually orthogonal mate. Clearly it could happen that the choice of the first two was unfortunate. The same method was applied to the plane of order 12. Here the trial and error method failed to produce even two orthogonal squares. It may be worthwhile to remark that the construction of the plane and consequently the search for orthogonal latin squares does not require the assumption that the oval consists of the maximum number of points $n + 2$. However, if the plane does not include an oval consisting of $n + 2$ points the lines could not be classified into two categories only and this complicates the construction of the plane. Let us illustrate the method in the case n equals 10. It is easy to show that in this case the oval must consist of at least 6 points. However, the case of an oval of 6 points would be ignored since in this case every quadrangle would have to have collinear diagonals. On the other hand, a plane of order ten must be a non-Desarguessian and hence must [illegible] a non-degenerate quadrangle with noncollinear diagonals. Suppose that the plane contains a quadrangle with noncollinear diagonals and suppose that the plane contains an oval consisting of seven points then the 104 points of the plane which do not belong to the oval could be classified into three categories:

(i) points lying on 3 secants, 1 tangent 7 nonintersectors

(ii) " " " 2 " 3 " 6 "

(iii) " " " 1 " 5 " 5 "

Let us name the number of points in each category by x, y, z respectively. Clearly $x + y + z = 104$.

Counting the intersections of the secants and the tangents we get the further equations:

$$3x + y = 105$$

$$3y + 10z = 525$$

The unique solutions of this system of equations are $x = 20$, $y = 45$, $z = 39$. One could start the construction of the plane under the present assumption and investigate the possibilities of obtaining orthogonal latin squares in this way.

X. Code Construction of $O(n,t)$ Sets

Given an n-symbol alphabet, e.g., $1, 2, \ldots, n$, and a set of k-tuples of the n symbols, we denote the set of all k-tuples by $C_{k,n}$. This set may be thought of as a vector space or as a k-dimensional hypercube with edges of length n. Any subset of $C_{k,n}$ is denoted as a block code with a block length of k. The elements of the subset are denoted as code words. The number of symbols by which any two code words differ is called the Hamming distance. If any pair of code words in the subset differs by a Hamming difference of at least r, the block code is called a distance r code. A distance r code is called an $(r-1)/2$-error correcting code because fewer than $(r-1)/2$ changes leaves the word closer to its original form than to any other code word in the subset. For similar reasons, a distance r code has also been designated as an $(r-1)$-error-dectecting code.

In an interesting paper, Golomb and Posner [1964] discuss the relationships between a subset of n^2 code words and an $O(n,t)$ set and relate these to ideas developed from a consideration of a set of n^2 super rooks of power t on the n^{t+2} chessboard such that no two super rooks attack each other. The new concepts of rook domains and rook packing were found to be very useful in providing a geometrical view of the results.

Any subset of n^2 words from $C_{3,n}$ which forms a single-error-detecting code may be used to construct a latin square of order n as any pair of the triples differs by at least two symbols. Likewise, any subset of n^2 words from $C_{t+2,n}$ with a Hamming distance of $t+1$ may be utilized to construct an $O(n,t)$ set. These results are embodied in the following theorem (from Golomb and Posner [1964]):

Theorem X.8.1 The following three concepts are equivalent:

i) an $O(n,t)$ set

ii) A set of n^2 nonattacking super rooks of power t on the n^{t+2} board.

For even t, also the following, a set of n^2 super rooks of power $t/2$ on the n^{t+2} board such that no cell is attacked twice; that is, such that the rook domains are nonoverlapping.

iii) A distance $t+1$ code of block length $t+2$ with n^2 words from an n-symbol alphabet.

For those interested in code construction, reference may be made to Mann [1968] and Peterson [1961] and the literature citations therein. We shall merely illustrate the method of construction of an $O(n,t)$ set from n^2 words of length $t+2$ and Hamming distance $t+1$ through an example. Let $n = 3$ and $t = 2$. Then the $n^2 = 9$ code words with length 4 and Hamming distance 3 and the corresponding latin squares are:

0000	0111	0222
1012	1120	1201
2021	2102	2210

to produce

latin squares of order 3

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

and

	0	1	2
0	0	1	2
1	2	0	1
2	1	2	0

where the first symbol corresponds to row number, the second to column number, the third to symbols in the first latin square, and the fourth to symbols in the second latin square. The two latin squares form an $O(3,2)$ set. Note that any pair of the quadruples differs in at least three symbols.

The analogy of the above with many of the concepts from fractional replication and orthogonal arrays is immediately apparent. The equivalences of many of the results in these fields need to be systematically noted much in the same manner that Golomb and Posner [1964] note various equivalences among $O(n,t)$ sets, error-correcting codes, and n^2 nonattacking rooks on an n^{t+2} chessboard.

XI. Pairwise Balanced Design Construction of $O(n,t)$ Sets

Central to the constructions of orthogonal latin squares of Bose and Shrikhande [1959] and of Parker [1959, 1960] is the following which might be called a "Folk theorem," being credited to no specific investigator: From a set of t orthogonal latin squares of order n one may produce a set of n^2 ordered $(t+2)$-tuples on n^2 symbols such that each pair of distinct positions contains each ordered pair of symbols (exactly once); the converse construction can also be carried out. (Some, such as Bose, prefer to call the set of $(t+2)$-tuples an orthogonal array.) There is nothing difficult to prove in this construction. Two arbitrary positions in the $(t+2)$-tuples are identified with row and column indices in matrices, and each other position with entries in one of the matrices. The equivalence between orthogonality of latin squares and the conditions on the $(t+2)$-tuples is then fairly apparent.

Parker [1960] contributed the following to the construction of orthogonal latin squares. If there exists a pair of orthogonal latin squares of order m, then there exists a pair of orthogonal latin squares of order $3m+1$.

Let the $3m+1$ symbols be $X_1, \ldots, X_m$ and the residue classes (mod $2m+1$). Form the (latin square) array

$$\begin{array}{cccc} X_i & 0 & 1 & -1 \\ 0 & X_i & -1 & 1 \\ 1 & -1 & X_i & 0 \\ -1 & 1 & 0 & X_i \end{array} .$$

for each i, $1 \le i \le m$, each row is one of the ordered quadruples. In turn, the list of quadruples is built up by adding each integer (mod $2m + 1$) to all four positions at once, the X_i symbols being unchanged by the addition. The set of $4m(2m + 1)$ ordered quadruples just described contains in each pair of distinct positions exactly one occurrence of each ordered pair made up of an X_i and a residue class in either order, and of each ordered pair made up of two distinct residue classes. The required set of ordered quadruples is completed by adjoining: i) all ordered quadruples (j,j,j,j), $j = 0, \ldots, 2m$; ii) a set of ordered quadruples on the X_i symbols corresponding to a pair of orthogonal latin squares of order m guaranteed by the hypothesis to exist.

Bose and Shrikhande (1959, published 1959 and 1960 partly in a 3-author paper with Parker) developed a sequence of constructive theorems which led in steps to disproof of Euler's conjecture for all orders $4t + 2 > 6$. Their central theorem given here does not exhaust their methods, but virtually all their results rest on this theorem. We begin with a definition. A pairwise balanced design, $PB(n; k_1, \ldots, k_t)$ is a collection of subsets of a set of n elements, each subset having number of elements one of the k_i, and such that each pair of distinct elements in the set of n occurs in a unique subset of the PB. (Note: unlike in balanced incomplete block designs, the subsets of a PB are not restricted to have equal numbers of elements.) Now for the main theorem of Bose and Shrikhande. If a $PB(n; k_1, \ldots, k_t)$ exists, and for each i, $1 \le i \le t$, a set of m orthogonal latin squares of order k_i exists, then a set of $m - 1$ orthogonal latin squares of order n exists. Loosely speaking, the sets of ordered tuples for each subset of the PB are constructed and these fit together to form a set of ordered tuples for the full set of n elements. The decrease from m to $m-1$ orthogonal latin squares occurs because in fitting the pieces together to form the large set of ordered tuples, it is necessary that each set of ordered tuples formed from a subset of the PB include each $(i, i, \ldots, i)$, where i ranges over the elements of that subset. (It is sufficient that this condition be fulfilled in the construction. Thus the theorem might be stated in slightly stronger form: "<u>If</u> ... $1 \leq i \leq t$, <u>a set of</u> m <u>orthogonal latin squares of order</u> k_i <u>with a transversal exists, then a set of</u> m <u>orthogonal latin squares of order</u> n <u>exists</u>.") Now for a more nearly formal version of the proof. If there exists a set of m orthogonal latin squares of order n, then there exists a set of the appropriate sort of n^2 ordered $(m+1)$-tuples with each symbol repeated in an $(m+1)$-tuple $m+1$ times. (The condition mentioned is satisfied with $(m+2)$-tuples if the set of orthogonal latin squares has a transversal.) One need simply put together the ordered tuples on each subset of the PB in turn, subject to the important condition that within each subset of the PB, each tuple of repetitions of each symbol be included. Carrying this out on the alphabet of the symbols in each subset of the PB, one has the construction for the set of orthogonal latin squares in the conclusion: each ordered tuple of a repeated symbol among the n is used only once.

A representative and very interesting example (Bose and Shrikhande informed Parker that this was the first case of disproof of Euler's conjecture produced in their joint work at a blackboard) yields 5 mutually orthogonal latin squares of order 50 via the PB construction. One forms the affine plane of order 7, then adjoins exactly one ideal point on each line of one class of parallel lines. This yields a PB(50; 8,7). Since there exist 6 orthogonal latin squares of each order 8 and 7, there exist $6-1=5$ orthogonal latin squares of order 50.

There is a limitation on the Bose-Shrikhande PB construction. Aside from trivial PB designs, having a single subset of all elements, any PB has a subset with at most one more element than the square root of the number of elements in the large set. Thus other techniques are requisite to produce more than $\sqrt{n}$ orthogonal latin squares of order not a prime-power.

XII. Product Composition of O(n,t) Sets

About 70 years ago, for the first time, Tarry [1899] in his half-page note asserted that if there exists an O(a,2) set and if there exists an O(b,2) set then there exists an O(ab,2) set. He exhibited the following O(12,2) set, by composing two O(3,2) and O(4,2) sets, to demonstrate the truth of his assertion. Note that in the following square the set of first integers belong to one latin square and the set of second integers belong to the second latin square. No more description is given by Tarry.

2-3	1-1	3-2	8-12	7-10	4-11	11-6	10-4	12-5	5-9	4-7	6-8
3-1	2-2	1-3	9-10	8-11	7-12	12-4	11-5	10-6	6-7	5-8	4-9
1-2	3-3	2-1	7-11	0-12	8-10	10-5	12-6	11-4	4-8	6-9	5-7
11-9	10-7	12-8	5-6	4-4	6-5	2-12	1-10	3-11	8-3	7-1	9-2
12-7	11-8	10-9	6-4	5-5	4-6	3-10	2-11	1-12	9-1	8-2	7-3
10-8	12-9	11-7	4-5	6-6	5-4	1-11	3-12	2-10	7-2	9-3	8-1
5-12	4-10	6-11	11-3	10-1	12-2	8-9	7-7	9-8	2-6	1-4	3-5
6-10	5-11	4-12	12-1	11-2	10-3	9-7	8-8	7-9	3-4	2-5	1-6
4-11	6-12	5-10	10-2	12-3	11-1	7-8	9-9	8-7	1-5	3-6	2-4
8-6	7-4	9-5	2-9	1-7	3-8	5-3	4-1	6-2	11-12	10-10	12-11
9-4	8-5	7-6	3-7	2-8	1-9	6-1	5-2	4-3	12-10	11-11	10-12
7-5	9-6	8-4	1-8	3-9	2-7	4-2	6-3	5-1	10-11	12-12	11-10

Tarry did not observe any generalization of his method. Perhaps this was due to the fact that he, like so many other researchers, was only concerned with sets of type O(n,2). Probably he was not aware of the existence of a larger set.

About 23 years later MacNeish [1922] demonstrated:

1) The existence and a construction of an $O(n, n-1)$ set for n a prime or prime power integer.

2) A generalization of Tarry's procedure viz., if there exists an $O(a, r)$ set and if there exists an $O(b, r)$ set then there exists an $O(ab, r)$ set.

3) By a successive application of 1) and 2) he showed that if $n = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_t^{\alpha_t}$ is the prime-power decomposition of n then there exists an $O(n, r)$ set where $r = \min\{p_i^{\alpha_i} - 1, \; i = 1, 2, \ldots, t\}$.

MacNeish could not embed his $O(n, r)$ set generated in 3) in a larger set. This unsuccessful attempt, reinforced by Euler's conjecture, led MacNeish to prove (erroneously) geometrically that $O(n, t)$ sets do not exist for $t > r$, and therefore, as a confirmation of Euler's conjecture. The preceding argument of MacNeish is known as MacNeish's conjecture in the literature. By constructing an $O(21, 3)$ set Parker [1959] gave a counter example to MacNeish's conjecture. Later Bose, Shrikhande, and Parker [1960] completely demolished Euler's conjecture except for $n = 6$. It should be mentioned that MacNeish's conjecture has not been totally disproved yet. For instance, no one as yet as far as we know, has constructed an $O(15, 4)$ set (an $O(15, 3)$ set is given in section V for the first time) or an $O(20, 3)$ set. We believe that MacNeish should be given substantial credit for his non-erroneous contributions. It is to be regretted that MacNeish is often cited in the literature only for his false conjecture.

Even though Tarry and MacNeish did not attach any name to their procedure, it is not difficult to see that it is the method of Kronecker product of matrices. Therefore, we can state, more formally, their results as follows:

Theorem (Tarry-MacNeish). If $\{A_1, A_2, \ldots, A_r\}$ is an $O(n,r)$ set and if $\{B_1, B_2, \ldots, B_r\}$ is an $O(m,r)$ set, then $\{A_1 \otimes B_1, A_2 \otimes B_2, \ldots, A_r \otimes B_r\}$, where $\otimes$ denotes the Kronecker product operation of matrices, is an $O(nm, r)$ set.

The preceding arguments clearly support the choice of the title for this section and is in contrast to the choice of the name for the procedure given in section XIII.

XIII. Sum Composition Construction of O(n,t) Sets

XIII.1. Introduction. Perhaps one of the most useful techniques for the construction of combinatorial systems is the method of composition. To mention some, here are few well-known examples: 1) If there exists a set of t orthogonal latin squares of order n_1 and if there exists a set of t orthogonal latin squares of order n_2, then there exists a set of t orthogonal latin squares of order n_1n_2. 2) If there are Steiner triple systems of order v_1 and v_2, there is a Steiner triple system of order $v = v_1v_2$. 3) If H_1 and H_2 are two Hadamard matrices of order n_1 and n_2 respectively, then the Kronecker product of H_1 and H_2 is a Hadamard matrix of order n_1n_2. 4) If Room squares of order n_1 and n_2 exist, then a Room square of order n_1n_2 exists. 5) If BIB (v_1, k, λ_1) and BIB (v_2, k, λ_2) exist and if $f(\lambda_2 v_2^2) \geq k$, then BIB $(v_1v_2, k, \lambda_1\lambda_2)$ exists where $f(\lambda_2 v_2^2)$ denotes the maximum number of constraints which are possible in an orthogonal array of size $\lambda_2 v_2^2$, with v_2 levels, strength 2, and index λ_2. 6) As a final example, the existence of orthogonal arrays $(\lambda_i v_i^t, q_i, v_i, t)$, $i = 1, 2, \ldots, r$ implies the existence of the orthogonal array $(\lambda v^t, q, v, t)$, where $\lambda = \lambda_1\lambda_2\ldots\lambda_r$, $v = v_1v_2\ldots v_r$, and $q = \min(q_1, q_2, \ldots, q_r)$.

The reader will note that each of the above examples involved a product type composition. The method that we will describe utilizes a sum type composition, by means of which one can possibly construct sets of orthogonal latin squares for all $n \geq 10$.

XIII.2. Definitions. In the sequel by an $O(n,t)$ set we mean a set of t mutually orthogonal latin squares of order n.

a) A <u>transversal</u> (directrix) of a latin square L of order n on an n-set Σ is a collection of n cells such that the entries of these cells exhaust the set Σ and every row and column of L is represented in this collection. Two transversals are said to be parallel if they have no cell in common.

b) A collection of n cells is said to form a <u>common transversal</u> for an $O(n, t)$ set if the collection is a transversal for each of these t latin squares. Similarly, two common transversals are said to be parallel if they have no cell in common.

<u>Example</u>. The underlined and parenthesized cells form two parallel common transversals for the following $O(4, 2)$ set.

$$\left\{\begin{matrix} 1 & 2 & (3) & \underline{4} \\ (2) & \underline{1} & 4 & 3 \\ \underline{3} & (4) & 1 & 2 \\ 4 & 3 & \underline{2} & (1) \end{matrix} \qquad \begin{matrix} 1 & 2 & (3) & \underline{4} \\ (4) & \underline{3} & 2 & 1 \\ \underline{2} & (1) & 4 & 3 \\ 3 & 4 & \underline{1} & (2) \end{matrix}\right\}$$

XIII. 3. <u>Composing Two Latin Squares of Order</u> n_1 <u>and</u> n_2

A very natural question in the theory of latin squares is the following: Given two latin squares L_1 and L_2 of order n_1 and n_2 $(n_1 \geq n_2)$ respectively. In how many ways can one compose L_1 and L_2 in order to obtain a latin square L_3 of order m, where m is a function of n_1 and n_2 only? This question can be partially answered as follows. First, it is well-known that the Kronecker product $L_3 = L_1 \otimes L_2$ is a latin square of order $m = n_1 n_2$ irrespective of the combinatorial structure of L_1 and L_2. Secondly, we show that if L_1 has a certain combinatorial structure, then one can construct a latin square L of order $n = n_1 + n_2$. Naturally enough we call this procedure a "method of sum composition".

Even though our method of sum composition does not work for all pairs of latin squares, it has an immediate application in the construction of orthogonal latin squares including those of order $4t + 2$, $t \geq 2$. We emphasize that the combinatorial structure of orthogonal latin squares constructed by the method of sum composition is completely different from those of known orthogonal latin squares in the literature. Therefore, it is worthwhile to study these squares for the purpose of constructing new finite projective planes.

We shall now describe the method of "sum composition". Let L_1 and L_2 be two latin squares of order n_1 and n_2, $n_1 \geq n_2$, on two non-intersecting sets $\Sigma_1 = \{a_1, a_2, \ldots, a_{n_1}\}$ and $\Sigma_2 = \{b_1, b_2, \ldots, b_{n_2}\}$ respectively. If L_1 has n_2 parallel transversals then we can compose L_1 with L_2 to obtain a latin square L of order $n = n_1 + n_2$. Note that for any pair (n_1, n_2), there exists L_1 and L_2 with the above requirement, except for $(2,1)$, $(2,2)$, $(6,5)$ and $(6,6)$.

To produce L put L_1 and L_2 in the upper left and lower right corner respectively. Call the resulting square C_1, which looks as follows:

$$C_1 = \begin{array}{|c|c|}\hline L_1 & \\ \hline & L_2 \\ \hline\end{array}$$

Name the n_2 transversals of L_1 in any manner from 1 to n_2. Now fill the cell $(i, n_1 + k)$, $k = 1, 2, \ldots, n_2$, with that element of transversal k which appears in row i, $i = 1, 2, \ldots, n_1$. Fill also the cell $(n_1 + k, j)$, $k = 1, 2, \ldots, n_2$, with that element of transversal k which appears in column j, $j = 1,2,\ldots,n_1$. Call the resulting square C_2. Now every entry of C_2 is occupied with an element either from Σ_1 or Σ_2, but C_2 is obviously not a latin square on $\Sigma_1 \cup \Sigma_2$. However, if we replace each of the n_1 entries of transversal k with b_k, it is easily verified that the resulting square which we call L is a latin square of order n on $\Sigma_1 \cup \Sigma_2$.

The procedure described for filling the first n_1 entries of the row (column) $n_1 + k$ with the corresponding entries of transversal k is, naturally enough, called <u>the projection of transversal k on the first n_1 entries of row (column)</u> $n_1 + k$.

We shall now elucidate the above procedure via an example. Let $\Sigma_1 = \{1,2,3,4,5\}$, $\Sigma_2 = \{6,7,8\}$,

$$L_1 = \begin{array}{ccccc|c} 1 & 2 & 3 & 4 & 5 & 1 \\ 5 & 1 & 2 & 3 & 4 & 2 \quad 3 \\ 4 & 5 & 1 & 2 & 3 & \\ 3 & 4 & 5 & 1 & 2 & \\ 2 & 3 & 4 & 5 & 1 & \end{array} \quad \text{and} \quad L_2 = \begin{array}{ccc} 6 & 7 & 8 \\ 7 & 8 & 6 \\ 8 & 6 & 7 \end{array}.$$

Note that the cells on the same curve in L_1 form a transversal.

$$C_1 = \begin{array}{|ccccc|ccc|} \hline 1 & 2 & 3 & 4 & 5 & & & \\ 5 & 1 & 2 & 3 & 4 & & & \\ 4 & 5 & 1 & 2 & 3 & & & \\ 3 & 4 & 5 & 1 & 2 & & & \\ 2 & 3 & 4 & 5 & 1 & & & \\ \hline & & & & & 6 & 7 & 8 \\ & & & & & 7 & 8 & 6 \\ & & & & & 8 & 6 & 7 \\ \hline \end{array} \quad \text{and} \quad C_2 \quad \begin{array}{|ccccc|ccc|} \hline 1 & 2 & 3 & 4 & 5 & 1 & 2 & 3 \\ 5 & 1 & 2 & 3 & 4 & 4 & 5 & 1 \\ 4 & 5 & 1 & 2 & 3 & 2 & 3 & 4 \\ 3 & 4 & 5 & 1 & 2 & 5 & 1 & 2 \\ 2 & 3 & 4 & 5 & 1 & 3 & 4 & 5 \\ \hline 1 & 3 & 5 & 2 & 4 & 6 & 7 & 8 \\ 5 & 2 & 4 & 1 & 3 & 7 & 8 & 6 \\ 4 & 1 & 3 & 5 & 2 & 8 & 6 & 7 \\ \hline \end{array}$$

And finally

$$L = \begin{array}{|ccccc|ccc|} 6 & 7 & 8 & 4 & 5 & 1 & 2 & 3 \\ 7 & 8 & 2 & 3 & 6 & 4 & 5 & 1 \\ 8 & 5 & 1 & 6 & 7 & 2 & 3 & 4 \\ 3 & 4 & 6 & 7 & 8 & 5 & 1 & 2 \\ 2 & 6 & 7 & 8 & 1 & 3 & 4 & 5 \\ \hline 1 & 3 & 5 & 2 & 4 & 6 & 7 & 8 \\ 5 & 2 & 4 & 1 & 3 & 7 & 8 & 6 \\ 4 & 1 & 3 & 5 & 2 & 8 & 6 & 7 \end{array}$$

which is a latin square of order 8 on $\Sigma_1 \cup \Sigma_2 = \{1, 2, \ldots, 8\}$.

Remark. Note that it is by no means required that the projection of transversals on the rows and columns should have the same ordering. Indeed, for the fixed set of ordered n_2 transversals, we have $n_2!$ choices of projections on columns and $n_2!$ choices of projections on the rows. Hence we can generate at least $(n_2!)^2$ different latin squares of order $n = n_1 + n_2$ composing L_1 and L_2.

XIII. 4. Construction of $O(n, 2)$ Sets by Method of Sum Composition. In order to construct an $O(n, 2)$ set for $n = n_1 + n_2$, we require that $n_1 \geq 2n_2$ and there should exist an $O(n_2, 2)$ set, and an $O(n_1, 2)$ set with $2n_2$ parallel transversals. It is easy to show that any $n \geq 10$ can be decomposed in at least one way into $n_1 + n_2$ which fulfill the above requirements. We now present two theorems which state that for certain n one can construct an $O(n, 2)$ set by the method of sum composition.

Theorem XIII. 4. 1. *Let $n_1 = p^\alpha \geq 7$ for any odd prime p and positive integer α, excluding $n_1 = 13$. Then there exists an $O(n, 2)$ set which can be constructed by composition of two $O(n_1, 2)$ and $O(n_2, 2)$ sets for $n_2 = (n_1 - 1)/2$ and $n = n_1 + n_2$.*

We shall first give the method of construction and then a proof that the constructed set is an $O(n,2)$ set.

Construction. Let $B(r)$ be the $n_1 \times n_1$ square with element $r\alpha_i + \alpha_j$ in its (i,j) cell, $\alpha_i, \alpha_j, 0 \neq r$ in $GF(n_1)$, $i,j = 1,2,\ldots,n_1$. Then it is easy to see that $\{B(1), B(x), B(y)\}$, $y = x^{-1}$, $x \neq 1$, is an $O(n_1,3)$ set. Consider the n_1 cells in $B(1)$ with $\alpha_i + \alpha_j = k$ a fixed element in $GF(n_1)$. Then the corresponding cells in $B(x)$ and $B(y)$ form a common transversal for the set $\{B(x), B(y)\}$. Name this common transversal by k. It is then obvious that two common transversals k_1 and k_2, $k_1 \neq k_2$ are parallel and hence $\{B(x), B(y)\}$ has n_1 common parallel transversals. Now let $\{A_1, A_2\}$ be any $O(n_2,2)$ set, which always exists, on a set Ω non-intersecting with $GF(n_1)$. For any λ in $GF(n_1)$ we can find $(n_1-1)/2$ pairs of distinct elements belonging to $GF(n_1)$ such that the sum of the two elements of each pair is equal to λ. Let $\{S\}$ and $\{T\}$ denote the collection of the first and the second elements of these $(n_1-1)/2$ pairs respectively. Note that for a fixed λ the set $\{S\}$ can be constructed in $(n_1-1)(n_1-3)\ldots1$ distinct ways. Now fix λ and let L_1 denote any of the $(n_2!)^2$ latin squares that can be generated by the sum composition of $L(x)$ and A_1 using transversals determined by the n_2 elements of $\{S\}$. Let L_2 be the latin square derived from the composition of $L(y)$ and A_2 using the n_2 transversals determined by the elements of $\{T\}$ and the following projection rule: Project transversals t_i, $i = 1,2,\ldots,n_2$ on the row (column) which upon superposition of L_2 on L_1 this row (column) should coincide with the row (column) stemmed from the transversal $\lambda - t_i$. Shortly we shall prove that $\{L_1, L_2\}$ forms on $O(n,2)$ set.

The preceding arguments shows that $\{L_1, L_2\}$ can be constructed non-isomorphically in at least $(n_1-3)(n_2!)^2[n_1(n_1-1)(n_1-3)\ldots 1]$ ways. For instance in the case of $n_1 = 7$, there is at least 12096 non-isomorphic pairs of orthogonal latin squares of order 10. Therefore, Euler has been wrong in his conjecture by a very wide margin.

Note that we can construct infinitely many pairs of orthogonal latin squares of order $4t + 2$ by the method of theorem XIII. 4.1. For $p \equiv 7 \bmod 8$ and α odd $p^\alpha = (8t + 5)/3)$. Hence $n_1 + n_2 = 4t + 2$.

Proof: The constructional procedure clearly reveals that:

A. L_1 and L_2 are latin squares of order n on $GF(n_1) \cup \Omega$.

B. Upon superposition of L_1 on L_2 the following are true:

b_1. Every element of Ω appears with every other element of Ω.

b_2. Every element of Ω appears with every element of $GF(n_1)$.

b_3. Every element of $GF(n_1)$ appears with every element of Ω.

Therefore, all we have to prove is that every element of $GF(n_1)$ appears with every other element of $GF(n_1)$. To prove this recall that $B(x)$ is orthogonal to $B(y)$. However, since we removed the n_2 transversals from $B(x)$ determined by the n_2 elements of $\{S\}$ and n_2 transversals from $B(y)$ determined by the n_2 elements of $\{T\}$ therefore the following $2n_2n_1$ pairs have been lost.

$$(x\alpha_i + \alpha_j, y\alpha_i + \alpha_j) \text{ with } \alpha_i + \alpha_j = \gamma \text{ for any } \gamma \in GF(n_1), \ \gamma \neq \lambda .$$

We claim that the given projection rules guarantee the capture of these lost pairs by the $2n_2n_1$ bordered cells. To show this note that the superposition of the projected transversal s from $B(x)$ on the projected transversal $t = \lambda - s$ from $B(y)$ will capture the n_1 pairs.

$$(x\alpha_i + \alpha_j,\ y\alpha_i + \alpha_j) \text{ with } \alpha_i + \alpha_j = k = [y(\lambda - s) + s]/(1+y)$$

if these transversals have been projected on row border and n_1 pairs

$$(x\alpha_i + \alpha_j,\ y\alpha_i + \alpha_j) \text{ with } \alpha_i + \alpha_j = k = [s(y-1) + (s-\lambda)(x-1)]/(y-x)$$

if these transversals have been projected on column border. Now because $k + k' = \lambda$ and if $s_1 \neq s_2$ then $k_1 \neq k_2$ and $k_1' \neq k_2'$ hence the $2n_2n_1$ pairs which have been resulted from the projection of transversals determined by $\{S\}$ and $\{T\}$ will jointly capture the $2n_2n_1$ lost pairs and thus a proof.

We shall now clarify the above constructional procedure by an example.

Example. Let $n_1 = 7$, $GF(7) = \{0,1,2,\ldots,6\}$. Then for $x = 2$, $y = x^{-1} = 4$ we have

$\{B(1), B(2), B(4)\} =$

$$\begin{matrix} 0&1&2&3&4&5&6\\ 1&2&3&4&5&6&0\\ 2&3&4&5&6&0&1\\ 3&4&5&6&0&1&2\\ 4&5&6&0&1&2&3\\ 5&6&0&1&2&3&4\\ 6&0&1&2&3&4&5 \end{matrix} \qquad \begin{matrix} 0&1&2&3&4&5&6\\ 2&3&4&5&6&0&1\\ 4&5&6&0&1&2&3\\ 6&0&1&2&3&4&5\\ 1&2&3&4&5&6&0\\ 3&4&5&6&0&1&2\\ 5&6&0&1&2&3&4 \end{matrix} \qquad \begin{matrix} 0&1&2&3&4&5&6\\ 4&5&6&0&1&2&3\\ 1&2&3&4&5&6&0\\ 5&6&0&1&2&3&4\\ 2&3&4&5&6&0&1\\ 6&0&1&2&3&4&5\\ 3&4&5&6&0&1&2 \end{matrix}$$

For $n_2 = (n_1-1)/2 = 3$ let $\Omega_2 = \{7,8,9\}$ and

$\{A_1, A_2\} = \begin{matrix} 7&8&9\\ 8&9&7\\ 9&7&8 \end{matrix}, \begin{matrix} 7&8&9\\ 9&7&8\\ 8&9&7 \end{matrix}$. Finally for $\lambda = 0$, $\{S\} = \{1,2,3\}$ and $\{T\} = \{6,5,4\}$ we have $\{L_1, L_2\} =$

$$
\begin{array}{ccccccc|ccc}
0 & 7 & 8 & 9 & 4 & 5 & 6 & 1 & 2 & 3 \\
7 & 8 & 9 & 5 & 6 & 0 & 1 & 2 & 3 & 4 \\
8 & 9 & 6 & 0 & 1 & 2 & 7 & 3 & 4 & 5 \\
9 & 0 & 1 & 2 & 3 & 7 & 8 & 4 & 5 & 6 \\
1 & 2 & 3 & 4 & 7 & 8 & 9 & 5 & 6 & 0 \\
3 & 4 & 5 & 7 & 8 & 9 & 2 & 6 & 0 & 1 \\
5 & 6 & 7 & 8 & 9 & 3 & 4 & 0 & 1 & 2 \\
\hline
2 & 1 & 0 & 6 & 5 & 4 & 3 & 7 & 8 & 9 \\
4 & 3 & 2 & 1 & 0 & 6 & 5 & 8 & 9 & 7 \\
6 & 5 & 4 & 3 & 2 & 1 & 0 & 9 & 7 & 8
\end{array}
\qquad
\begin{array}{ccccccc|ccc}
0 & 1 & 2 & 3 & 7 & 8 & 9 & 6 & 5 & 4 \\
4 & 5 & 6 & 7 & 8 & 9 & 3 & 2 & 1 & 0 \\
1 & 2 & 7 & 8 & 9 & 6 & 0 & 5 & 4 & 3 \\
5 & 7 & 8 & 9 & 2 & 3 & 4 & 1 & 0 & 6 \\
7 & 8 & 9 & 5 & 6 & 0 & 1 & 4 & 3 & 2 \\
8 & 9 & 1 & 2 & 3 & 4 & 7 & 0 & 6 & 5 \\
9 & 4 & 5 & 6 & 0 & 7 & 8 & 3 & 2 & 1 \\
\hline
3 & 0 & 4 & 1 & 5 & 2 & 6 & 7 & 8 & 9 \\
6 & 3 & 0 & 4 & 1 & 5 & 2 & 9 & 7 & 8 \\
2 & 6 & 3 & 0 & 4 & 1 & 5 & 8 & 9 & 7
\end{array}
$$

the reader can easily verify that $\{L_1, L_2\}$ is an $O(10,2)$ set.

Remarks.

1) The method of theorem XIII.4.1 fails for $n_1 = 13$ only because there is no $O(6,2)$ set. Otherwise, there will be no orthogonality contradiction on the other parts of L_1 and L_2 with their 6×6 lower right square missing.

2) In the case of $n_1 = 7$, if we let $\{S\} = \{0,1,3\}$ and $\{T\} = \{2,4,5\}$ then the requirement $y = x^{-1}$ is not necessary. However then we do not have a unified projection rule for the formation of L_2 as was provided for the case $y = x^{-1}$ by theorem XIII.4.1. To give the complete list of solutions let (a_1,a_2,a_3) and (b_1,b_2,b_3) be any two permutations of the set $\{8,9,10\}$. If we project transversals $(0,1,3)$ on the rows (a_1,a_2,a_3) and columns (b_1,b_2,b_3) in the formation of L_1, then the following table indicates what permutation of transversals $\{2,4,5\}$ should be projected on the rows (a_1,a_2,a_3) and columns (b_1,b_2,b_3) in the formation of L_2. Obviously these permutations will be a function of the pair (x,y).

Pair (x,y)	Rows a_1,a_2,a_3	Columns b_1,b_2,b_3
(2,3)	4, 2, 5	4, 2, 5
(2,3)	2, 5, 4	2, 5, 4
(2,4)	2, 5, 4	4, 2, 5
(2,5)	4, 2, 5	4, 2, 5
(2,6)	2, 5, 4	2, 5, 4
(3,4)	2, 5, 4	2, 5, 4
(3,5)	2, 5, 4	4, 2, 5
(3,5)	4, 2, 5	5, 4, 2
(3,5)	4, 2, 5	2, 5, 4
(3,5)	5, 4, 2	2, 5, 4
(3,6)	4, 2, 5	2, 5, 4
(3,6)	5, 4, 2	4, 2, 5
(4,5)	2, 5, 4	2, 5, 4
(4,6)	5 4, 2	4, 2, 5
(4,6)	2, 5, 4	2, 5, 4
(4,6)	5, 4, 2	5, 4, 2

(This table is by no means exhaustive.)

The reader may note that whenever $y = x^{-1}$ in the above table the given solution(s) are different from the one provided by the method of theorem XIII. 4.1.

Thus we can conclude that any pair of orthogonal latin squares of order 7 based on the GF(7) can be composed with a pair of orthogonal latin squares of order 3 and make a pair of orthogonal latin squares of order 10. In addition, since we have six choices for (a_1, a_2, a_3) and (b_1, b_2, b_3) hence from every line in the above table we can produce 36 non-isomorphic $O(10,2)$ sets or $16 \times 36 = 576$ sets for the entire table. Since all these pairs are non-isomorphic with all previous pairs, produced by theorem XIII.4.1, thus by the method of sum composition one can at least produce 12,672 non-isomorphic $O(10,2)$ sets.

We believe that for other values of n_1 there are sets of $\{S\}$ and $\{T\}$ together with proper projections which makes the restriction $y = x^{-1}$ unnecessary.

Theorem XIII.4.2. Let $n_1 = 2^\alpha \geq 8$ for any positive integer α. Then there exists an $O(n,2)$ set which can be constructed by composition of two $O(n_1,2)$ and $O(n_2,2)$ sets for $n_2 = n_1/2$ and $n = n_1 + n_2$.

We shall here give only the method of construction. A similar argument as in theorem XIII.4.1 will show that the constructed set is an $O(n,2)$ set.

Construction. In a similar fashion as in theorem XIII.4.1 construct the set $\{B(1), B(x), B(y)\}$ over $GF(2^\alpha)$. Let also $\{A_1, A_2\}$ be any $O(n_2,2)$ set, which always exists, on a set Ω non-intersecting with $GF(2^\alpha)$. For any $\lambda \neq 0$ in $GF(2^\alpha)$ we can find $n_1/2$ pairs of distinct elements belonging to $GF(2^\alpha)$ such that the sum of the two elements of each pair is equal to λ. Let $\{S\}$ and $\{T\}$ denote the collection of the first and the second elements of these $n_1/2$ pairs respectively. Note that for a fixed λ the set $\{S\}$ can be constructed in $n_1(n_1-2)(n_1-4)\ldots 1$ distinct ways. Now form L_1 from the sum composition of $B(x)$ and A_1 and L_2 from the sum composition of $B(y)$ and A_2 using the same projection rule as given in theorem XIII.4.1. Now $\{L_1, L_2\}$ is an $O(n,2)$ set.

__Example.__ Let $n = 8$, $GF(8) = \{0,1,2,\ldots,7\}$ with the following addition $(+)$ and multiplication $(\times)$ tables:

+	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	0	6	4	3	7	2	5
2	2	6	0	7	5	4	1	3
3	3	4	7	0	1	6	5	2
4	4	3	5	1	0	2	7	6
5	5	7	4	6	2	0	3	1
6	6	2	1	5	7	3	0	4
7	7	5	3	2	6	1	4	0

×	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	3	4	5	6	7	1
3	0	3	4	5	6	7	1	2
4	0	4	5	6	7	1	2	3
5	0	5	6	7	1	2	3	4
6	0	6	7	1	2	3	4	5
7	0	7	1	2	3	4	5	6

Then for $x = 2$, $y = x^{-1} = 7$ we have

$\{B(1),\ B(2),\ B(7)\} =$

$$
\begin{array}{cccccccc}
0&1&2&3&4&5&6&7\\
1&0&6&4&3&7&2&5\\
2&6&0&7&5&4&1&3\\
3&4&7&0&1&6&5&2\\
4&3&5&1&0&2&7&6\\
5&7&4&6&2&0&3&1\\
6&2&1&5&7&3&0&4\\
7&5&3&2&6&1&4&0
\end{array}
\qquad
\begin{array}{cccccccc}
0&1&2&3&4&5&6&7\\
2&6&0&7&5&4&1&3\\
3&4&7&0&1&6&5&2\\
4&3&5&1&0&2&7&6\\
5&7&4&6&2&0&3&1\\
6&2&1&5&7&3&0&4\\
7&5&3&2&6&1&4&0\\
1&0&6&4&3&7&2&5
\end{array}
\qquad
\begin{array}{cccccccc}
0&1&2&3&4&5&6&7\\
7&5&3&2&6&1&4&0\\
1&0&6&4&3&7&2&5\\
2&6&0&7&5&4&1&3\\
3&4&7&0&1&6&5&2\\
4&3&5&1&0&2&7&6\\
5&7&4&6&2&0&3&1\\
6&2&1&5&7&3&0&4
\end{array}
$$

For $n_2 = n_1/2 = 4$ let $\Omega = \{A, B, C, D\}$ and

$$
\{A_1, A_2\} =
\begin{array}{cccc}
A&B&C&D\\
B&A&D&C\\
C&D&A&B\\
D&C&B&A
\end{array}
\qquad
\begin{array}{cccc}
A&B&C&D\\
D&C&B&A\\
B&A&D&C\\
C&D&A&B
\end{array}
$$

[illegible] $S = \{0, 1, \ldots 4$ [illegible] $6, 2\}$ we have $\{L_1, L_2\} =$

$$
\begin{array}{cccccccc|cccc}
A&B&2&C&D&5&6&7&0&1&3&4\\
B&A&0&D&C&4&1&3&6&2&5&7\\
3&4&A&0&1&D&B&C&7&5&2&6\\
C&D&5&A&B&2&7&6&1&0&4&3\\
D&C&4&B&A&0&3&1&2&6&7&5\\
6&2&D&5&7&A&C&B&3&4&0&1\\
7&5&B&2&6&C&A&D&4&3&1&0\\
1&0&C&4&3&B&D&A&5&7&6&2\\
\hline
0&6&7&1&2&3&4&5&A&B&C&D\\
2&1&3&6&0&7&5&4&B&A&D&C\\
4&7&6&3&5&1&0&2&C&D&A&B\\
5&3&1&7&4&6&2&0&D&C&B&A
\end{array}
\qquad
\begin{array}{cccccccc|cccc}
0&1&D&3&4&A&C&B&5&7&6&2\\
7&5&C&2&6&B&D&A&0&1&3&4\\
D&C&6&B&A&7&2&5&3&4&0&1\\
2&6&B&7&5&C&A&D&1&0&4&3\\
3&4&A&0&1&D&B&C&7&5&2&6\\
A&B&5&C&D&2&7&6&4&3&1&0\\
C&D&4&A&B&0&3&1&6&2&5&7\\
B&A&1&D&C&3&0&4&2&6&7&5\\
\hline
4&2&7&6&3&5&1&0&A&B&C&D\\
6&3&0&4&2&1&5&7&D&C&B&A\\
5&0&3&1&7&4&6&2&B&A&D&C\\
1&7&2&5&0&6&4&3&C&D&A&B
\end{array}
$$

which is an $O(12, 2)$ set.

__Discussion.__ The necessary requirements for the construction of an $O(n,t)$ set, $n = n_1 + n_2$, $t < n_2$, by the method of sum composition are: The existence of an $O(n_1,t)$ set, $n_1 \geq tn_2$, with at least tn_2 common parallel transversals, and an $O(n_2,t)$ set. These conditions are obviously satisfied whenever n_1 and n_2 are prime powers.

While for some values of n there exists only a unique decomposition fulfilling the above requirements, for infinitely many other values of n there are abundant such decompositions.

It seems that if there exists an $O(n_2,2)$ set and if $n = n_1 + n_2$, $n_1 \geq 2n_2$ then one can construct an $O(n,2)$ set by the method of sum composition if n_1 is a prime power. To support this observation and shed some more light on the method of sum composition we present in subsequent pages some highlights of the results which we hope to complete and submit for publication shortly.

In the following for each given decomposition of n we exhibit an $O(n,2)$ set which has been derived by the method of sum composition. We shall represent the pairs in a form that the curious reader can easily reconstruct the original sets. Hereafter the notation $L_1 \perp L_2$ means that L_1 is orthogonal to L_2 .

1) $12 = 9 + 3$

$$
\begin{array}{ccccccccc|ccc}
A & B & C & 4 & 5 & 6 & 7 & 8 & 9 & 1 & 2 & 3 \\
B & C & A & 1 & 2 & 3 & 4 & 5 & 6 & 9 & 7 & 8 \\
C & A & B & 7 & 8 & 9 & 1 & 2 & 3 & 5 & 6 & 4 \\
2 & 3 & 1 & 5 & 6 & 4 & A & B & C & 8 & 9 & 7 \\
8 & 9 & 7 & 2 & 3 & 1 & B & C & A & 4 & 5 & 6 \\
5 & 6 & 4 & 8 & 9 & 7 & C & A & B & 3 & 1 & 2 \\
3 & 1 & 2 & A & B & C & 9 & 7 & 8 & 6 & 4 & 5 \\
9 & 7 & 8 & B & C & A & 6 & 4 & 5 & 2 & 3 & 1 \\
6 & 4 & 5 & C & A & B & 3 & 1 & 2 & 7 & 8 & 9 \\
\hline
1 & 5 & 9 & 6 & 7 & 2 & 8 & 3 & 4 & A & B & C \\
7 & 2 & 6 & 3 & 4 & 8 & 5 & 9 & 1 & B & C & A \\
4 & 8 & 3 & 9 & 1 & 5 & 2 & 6 & 7 & C & A & B
\end{array}
\quad \perp \quad
\begin{array}{ccccccccc|ccc}
1 & 2 & 3 & 4 & 5 & 6 & A & B & C & 8 & 9 & 7 \\
9 & 7 & 8 & 3 & 1 & 2 & B & C & A & 6 & 4 & 5 \\
5 & 6 & 4 & 8 & 9 & 7 & C & A & B & 1 & 2 & 3 \\
6 & 4 & 5 & A & B & C & 3 & 1 & 2 & 7 & 8 & 9 \\
2 & 3 & 1 & B & C & A & 8 & 9 & 7 & 5 & 6 & 4 \\
7 & 8 & 9 & C & A & B & 4 & 5 & 6 & 3 & 1 & 2 \\
A & B & C & 2 & 3 & 1 & 5 & 6 & 4 & 9 & 7 & 8 \\
B & C & A & 7 & 8 & 9 & 1 & 2 & 3 & 4 & 5 & 6 \\
C & A & B & 6 & 4 & 5 & 9 & 7 & 8 & 2 & 3 & 1 \\
\hline
4 & 9 & 2 & 5 & 7 & 3 & 6 & 8 & 1 & A & B & C \\
3 & 5 & 7 & 1 & 6 & 8 & 2 & 4 & 9 & C & A & B \\
8 & 1 & 6 & 9 & 2 & 4 & 7 & 3 & 5 & B & C & A
\end{array}
$$

2) $14 = 11 + 3$, the only decomposition which fulfills the necessary requirements.

A	B	C	3	4	5	6	7	8	9	10	0	1	2
B	C	9	10	0	1	2	3	4	5	A	6	7	8
C	4	5	6	7	8	9	10	0	A	B	1	2	3
10	0	1	2	3	4	5	6	A	B	C	7	8	9
6	7	8	9	10	0	1	A	B	C	5	2	3	4
2	3	4	5	6	7	A	B	C	0	1	8	9	10
9	10	0	1	2	A	B	C	6	7	8	3	4	5
5	6	7	8	A	B	C	1	2	3	4	9	10	0
1	2	3	A	B	C	7	8	9	10	0	4	5	6
8	9	A	B	C	2	3	4	5	6	7	10	0	1
4	A	B	C	8	9	10	0	1	2	3	5	6	7
0	5	10	4	9	3	8	2	7	1	6	A	B	C
7	1	6	0	5	10	4	9	3	8	2	B	C	A
3	8	2	7	1	6	0	5	10	4	9	C	A	B

0	1	2	3	4	5	6	7	A	B	C	9	10	8
8	9	10	0	1	2	3	A	B	C	7	5	6	4
5	6	7	8	1	10	A	B	C	3	4	1	2	0
2	3	4	5	6	A	B	C	10	0	1	8	9	7
10	0	1	2	A	B	C	6	7	8	9	4	5	3
7	8	9	A	B	C	2	3	4	5	6	0	1	10
4	5	A	B	C	9	10	0	1	2	3	7	8	6
1	A	B	C	5	6	7	8	9	10	0	3	4	2
A	B	C	1	2	3	4	5	6	7	8	10	0	9
B	C	8	9	10	0	1	2	3	4	A	6	7	5
C	4	5	6	7	8	9	10	0	A	B	2	3	1
6	10	3	7	0	4	8	1	5	9	2	A	B	C
3	7	0	4	8	1	5	9	2	6	10	C	A	B
9	2	6	10	3	7	0	4	8	1	5	B	C	A

3) $15 = 12 + 3$, $15 = 11 + 4$ are the only decompositions which fulfill the necessary requirements. However, we consider here the latter decomposition since we can utilize the properties of Galois field GF(11).

A	B	C	D	4	5	6	7	8	9	10	0	1	2	3
B	C	D	5	6	7	8	9	10	0	A	1	2	3	4
C	D	6	7	8	9	10	0	1	A	B	2	3	4	5
D	7	8	9	10	0	1	2	A	B	C	3	4	5	6
8	9	10	0	1	2	3	A	B	C	D	4	5	6	0
10	0	1	2	3	4	A	B	C	D	9	5	6	0	1
1	2	3	4	5	A	B	C	D	10	0	6	7	1	2
3	4	5	6	A	B	C	D	0	1	2	7	8	2	3
5	6	7	A	B	C	D	1	2	3	4	8	9	3	4
7	8	A	B	C	D	2	3	4	5	6	9	10	4	5
9	A	B	C	D	3	4	5	6	7	9	10	0	5	6
0	10	9	8	7	6	5	4	3	2	1	A	B	C	D
2	1	0	10	9	8	7	6	5	4	3	B	A	D	C
4	3	2	1	0	10	9	8	7	6	5	C	D	A	B
6	5	4	3	2	1	0	10	9	8	7	D	C	B	A

0	1	2	3	4	5	A	B	C	D	10	8	6	9	7
6	7	8	9	10	A	B	C	D	4	5	2	0	3	1
1	2	3	4	A	B	C	D	9	10	0	7	5	8	6
7	8	9	A	B	C	D	3	4	5	6	1	10	2	0
2	3	A	B	C	D	8	9	10	0	1	6	4	7	5
8	A	B	C	D	2	3	4	5	6	7	0	9	1	10
A	B	C	D	7	8	9	10	0	1	2	5	3	6	4
B	C	D	1	2	3	4	5	6	7	A	10	8	0	9
C	D	6	7	8	9	10	0	1	A	B	4	2	5	3
D	0	1	2	3	4	5	6	A	B	C	9	7	10	8
5	6	7	8	9	10	0	A	B	C	D	3	1	4	2
9	4	10	5	0	6	1	7	2	8	3	A	B	C	D
10	5	0	7	1	7	2	8	3	9	4	D	C	B	A
3	9	4	10	5	0	6	1	7	2	8	B	A	D	C
4	10	5	0	6	1	7	2	8	3	9	C	D	A	B

4) $17 = 13 + 4$ and $17 = 12 + 5$ are the only decompositions which fulfill the necessary requirements.

The following pair is derived through the first decomposition.

A	B	C	D	4	5	6	7	8	9	10	11	12	0	1	2	3
B	C	D	8	9	10	11	12	0	1	2	3	A	4	5	6	7
C	D	12	0	1	2	3	4	5	6	7	A	B	8	9	10	11
D	3	4	5	6	7	8	9	10	11	A	B	C	12	0	1	2
7	8	9	10	11	12	0	1	2	A	B	C	D	3	4	5	6
12	0	1	2	3	4	5	6	A	B	C	D	11	7	8	9	10
4	5	6	7	8	9	10	A	B	C	D	2	3	11	12	0	1
9	10	11	12	0	1	A	B	C	D	6	7	8	2	3	4	5
1	2	3	4	5	A	B	C	D	10	11	12	0	6	7	8	9
6	7	8	9	A	B	C	D	1	2	3	4	5	10	11	12	0
11	12	0	A	B	C	D	5	6	7	8	9	10	1	2	3	4
3	4	A	B	C	D	9	10	11	12	0	1	2	5	6	7	8
8	A	B	C	D	0	1	2	3	4	5	6	7	9	10	11	12
0	9	5	1	10	6	2	11	7	3	12	8	4	A	B	C	D
5	1	10	6	2	11	7	3	12	8	4	0	9	B	A	D	C
10	6	2	11	7	3	12	8	4	0	9	5	1	C	D	A	B
2	11	7	3	12	8	4	0	9	5	1	10	6	D	C	B	A

0	1	2	3	4	5	6	7	A	B	C	D	12	9	8	11	10
8	9	10	11	12	0	1	A	B	C	D	6	7	3	2	5	4
3	4	5	6	7	8	A	B	C	D	0	1	2	10	9	12	7
11	12	0	1	2	A	B	C	D	7	8	9	10	4	3	6	5
6	7	8	9	A	B	C	D	1	2	3	4	5	11	10	0	12
1	2	3	A	B	C	D	8	9	10	11	12	0	5	4	7	6
9	10	A	B	C	D	2	3	4	5	6	7	8	12	11	1	0
4	A	B	C	D	9	10	11	12	0	1	2	3	6	5	8	7
A	B	C	D	3	4	5	6	7	8	9	10	11	0	12	2	1
B	C	D	10	11	12	0	1	2	3	4	5	A	7	6	9	8
C	D	4	5	6	7	8	9	10	11	12	A	B	1	0	3	2
D	11	12	0	1	2	3	4	5	6	A	B	C	8	7	10	9
5	6	7	8	9	10	11	12	0	A	B	C	D	2	1	4	3
7	0	6	12	5	11	4	10	3	9	2	8	1	A	B	C	D
12	5	11	4	10	3	9	2	8	1	7	0	6	D	C	B	A
10	3	9	2	8	1	7	0	6	12	5	11	4	B	A	D	C
2	8	1	7	0	6	12	5	11	4	10	3	9	C	D	A	B

5) We do not know whether there exists either an $O(14,2)$ set with 8 common parallel transversals or an $O(15,2)$ set with 6 common parallel transversals. Therefore the only decomposition of 18 which fulfills the necessary requirements is $18 = 13 + 5$. The following pair is constructed through this decomposition.

A	B	C	D	E	5	6	7	8	9	10	11	12	0	1	2	3	4
B	C	D	E	6	7	8	9	10	11	12	0	A	1	2	3	4	5
C	D	E	7	8	9	10	11	12	0	1	A	B	2	3	4	5	6
D	E	8	9	10	11	12	0	1	2	A	B	C	3	4	5	6	7
E	9	10	11	12	0	1	2	3	A	B	C	D	4	5	6	7	8
10	11	12	0	1	2	3	4	A	B	C	D	E	5	6	7	8	9
12	0	1	2	3	4	5	A	B	C	D	E	11	6	7	8	9	10
1	2	3	4	5	6	A	B	C	D	E	12	0	7	8	9	10	11
3	4	5	6	7	A	B	C	D	E	0	1	2	8	9	10	11	12
5	6	7	8	A	B	C	D	E	1	2	3	4	9	10	11	12	0
7	8	9	A	B	C	D	E	2	3	4	5	6	10	11	12	0	1
9	10	A	B	C	D	E	3	4	5	6	7	8	11	12	0	1	2
11	A	B	C	D	E	4	5	6	7	8	9	10	12	0	1	2	3
0	12	11	10	9	8	7	6	5	4	3	2	1	A	B	C	D	E
2	1	0	12	11	10	9	8	7	6	5	4	3	B	C	D	E	A
4	3	2	1	0	12	11	10	9	8	7	6	5	C	D	E	A	B
6	5	4	3	2	1	0	12	11	10	9	8	7	D	E	A	B	C
8	7	6	5	4	3	2	1	0	12	11	10	9	E	A	B	C	D

0	1	2	3	4	5	6	A	B	C	D	E	12	7	9	10	11	8
7	8	9	10	11	12	A	B	C	D	E	5	6	0	2	3	4	1
1	2	3	4	5	A	B	C	D	E	11	12	0	6	8	9	10	7
8	9	10	11	A	B	C	D	E	4	5	6	7	12	1	2	3	0
2	3	4	A	B	C	D	E	10	11	12	0	1	5	7	8	9	6
9	10	A	B	C	D	E	3	4	5	6	7	8	11	0	1	2	12
3	A	B	C	D	E	9	10	11	12	0	1	2	4	6	7	8	5
A	B	C	D	E	2	3	4	5	6	7	8	9	10	12	0	1	11
B	C	D	E	8	9	10	11	12	0	1	2	A	3	5	6	7	4
C	D	E	1	2	3	4	5	6	7	8	A	B	9	11	12	0	10
D	E	7	8	9	10	11	12	0	1	A	B	C	2	4	5	6	3
E	0	1	2	3	4	5	6	7	A	B	C	D	8	10	11	12	9
6	7	8	9	10	11	12	0	A	B	C	D	E	1	3	4	5	2
4	11	5	12	6	0	7	1	8	2	9	3	10	A	B	C	D	E
12	6	0	7	1	8	2	9	3	10	4	11	5	E	A	B	C	D
10	4	11	5	12	6	0	7	1	8	2	9	3	D	E	A	B	C
5	12	6	0	7	1	8	2	9	3	10	4	11	C	D	E	A	B
11	5	12	6	0	7	1	8	2	9	3	10	4	B	C	D	E	A

6) We do not know whether there exists either an $O(18,2)$ set with 8 common parallel transversals or an $O(15,2)$ set with 14 common parallel transversals. Therefore the only decomposition of 22 which fulfill the necessary requirements are $22 = 19 + 3$ and $22 = 17 + 5$.

a: $22 = 19 + 3$,

A	B	C	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	0	1	2
B	C	5	6	7	8	9	10	11	12	13	14	15	16	17	18	0	1	A	2	3	4
C	7	8	9	10	11	12	13	14	15	16	17	18	0	1	2	3	A	B	4	5	6
9	10	11	12	13	14	15	16	17	18	0	1	2	3	4	5	A	B	C	6	7	8
12	13	14	15	16	17	18	0	1	2	3	4	5	6	7	A	B	C	11	8	9	10
15	16	17	18	0	1	2	3	4	5	6	7	8	9	A	B	C	13	14	10	11	12
18	0	1	2	3	4	5	6	7	8	9	10	11	A	B	C	15	16	17	12	13	14
2	3	4	5	6	7	8	9	10	11	12	13	A	B	C	17	18	0	1	14	15	16
5	6	7	8	9	10	11	12	13	14	15	A	B	C	0	1	2	3	4	16	17	18
8	9	10	11	12	13	14	15	16	17	A	B	C	2	3	4	5	6	7	18	0	1
11	12	13	14	15	16	17	18	0	A	B	C	4	5	6	7	8	9	10	1	2	3
14	15	16	17	18	0	1	2	A	B	C	6	7	8	9	10	11	12	13	3	4	5
17	18	0	1	2	3	4	A	B	C	8	9	10	11	12	13	14	15	16	5	6	7
1	2	3	4	5	6	A	B	C	10	11	12	13	14	15	16	17	18	0	7	8	9
4	5	6	7	8	A	B	C	12	13	14	15	16	17	18	0	1	2	3	9	10	11
7	8	9	10	A	B	C	14	15	16	17	18	0	1	2	3	4	5	6	11	12	13
10	11	12	A	B	C	16	17	18	0	1	2	3	4	5	6	7	8	9	13	14	15
13	14	A	B	C	18	0	1	2	3	4	5	6	7	8	9	10	11	12	15	16	17
16	A	B	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	0
0	17	15	13	11	9	7	5	3	1	18	16	14	12	10	8	6	4	2	A	B	C
3	1	18	16	14	12	10	8	6	4	2	0	17	15	13	11	9	7	5	B	C	A
6	4	2	0	17	15	13	11	9	7	5	3	1	18	16	14	12	10	8	C	A	B

0	1	2	3	4	5	6	7	8	9	10	11	12	A	B	C	16	17	18	14	15	13
13	14	15	16	17	18	0	1	2	3	4	5	A	B	C	9	10	11	12	7	8	6
7	8	9	10	11	12	13	14	15	16	17	A	B	C	2	3	4	5	6	0	1	18
1	2	3	4	5	6	7	8	9	10	A	B	C	14	15	16	17	18	0	12	13	11
14	15	16	17	18	0	1	2	3	A	B	C	7	8	9	10	11	12	13	5	6	4
8	9	10	11	12	13	14	15	A	B	C	0	1	2	3	4	5	6	7	17	18	16
2	3	4	5	6	7	8	A	B	C	12	13	14	15	16	17	18	0	1	10	11	9
15	16	17	18	0	1	A	B	C	5	6	7	8	9	10	11	12	13	14	3	4	2
9	10	11	12	13	A	B	C	17	18	0	1	2	3	4	5	6	7	8	15	16	14
3	4	5	6	A	B	C	10	11	12	13	14	15	16	17	18	0	1	2	8	9	7
16	17	18	A	B	C	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	0
10	11	A	B	C	15	16	17	18	0	1	2	3	4	5	6	7	8	9	13	14	12
4	A	B	C	8	9	10	11	12	13	14	15	16	17	18	0	1	2	3	6	7	5
A	B	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18	0	17
B	C	13	14	15	16	17	18	0	1	2	3	4	5	6	7	8	9	A	11	12	10
C	6	7	8	9	10	11	12	13	14	15	16	17	18	0	1	2	A	B	4	5	3
18	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	A	B	C	16	17	15
12	13	14	15	16	17	18	0	1	2	3	4	5	6	7	A	B	C	11	9	10	8
6	7	8	9	10	11	12	13	14	15	16	17	18	0	A	B	C	4	5	2	3	1
11	18	6	13	1	8	15	3	10	17	5	12	0	7	14	2	9	16	4	A	B	C
5	12	0	7	14	2	9	16	4	11	18	6	13	1	8	15	3	10	17	C	A	B
17	5	12	0	7	14	2	9	16	4	11	18	6	13	1	8	15	3	10	B	C	A

b: 22 = 17 + 5,

A	B	C	D	E	5	6	7	8	9	10	11	12	13	14	15	16	0	1	2	3	4
B	C	D	E	6	7	8	9	10	11	12	13	14	15	16	0	A	1	2	3	4	5
C	D	E	7	8	9	10	11	12	13	14	15	16	0	1	A	B	2	3	4	5	6
D	E	8	9	10	11	12	13	14	15	16	0	1	2	A	B	C	3	4	5	6	7
E	9	10	11	12	13	14	15	16	0	1	2	3	A	B	C	D	4	5	6	7	8
10	11	12	13	14	15	16	0	1	2	3	4	A	B	C	D	E	5	6	7	8	9
12	13	14	15	16	0	1	2	3	4	5	A	B	C	D	E	11	6	7	8	9	10
14	15	16	0	1	2	3	4	5	6	A	B	C	D	E	12	13	7	8	9	10	11
16	0	1	2	3	4	5	6	7	A	B	C	D	E	13	14	15	8	9	10	11	12
1	2	3	4	5	6	7	8	A	B	C	D	E	14	15	16	0	9	10	11	12	13
3	4	5	6	7	8	9	A	B	C	D	E	15	16	0	1	2	10	11	12	13	14
5	6	7	8	9	10	A	B	C	D	E	16	0	1	2	3	4	11	12	13	14	15
7	8	9	10	11	A	B	C	D	E	0	1	2	3	4	5	6	12	13	14	15	16
9	10	11	12	A	B	C	D	E	1	2	3	4	5	6	7	8	13	14	15	16	0
11	12	13	A	B	C	D	E	2	3	4	5	6	7	8	9	10	14	15	16	0	1
13	14	A	B	C	D	E	3	4	5	6	7	8	9	10	11	12	15	16	0	1	2
15	A	B	C	D	E	4	5	6	7	8	9	10	11	12	13	14	16	0	1	2	3
0	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	A	B	C	D	E
2	1	0	16	15	14	13	12	11	10	9	8	7	6	5	4	3	B	C	D	E	A
4	3	2	1	0	16	15	14	13	12	11	10	9	8	7	6	5	C	D	E	A	B
6	5	4	3	2	1	0	16	15	14	13	12	11	10	9	8	7	D	E	A	B	C
8	7	6	5	4	3	2	1	0	16	15	14	13	12	11	10	9	E	A	B	C	D

0	1	2	3	4	5	6	7	8	A	B	C	D	E	14	15	16	10	11	12	13	9
9	10	11	12	13	14	15	16	A	B	C	D	E	5	6	7	8	1	2	3	4	0
1	2	3	4	5	6	7	A	B	C	D	E	13	14	15	16	0	9	10	11	12	8
10	11	12	13	14	15	A	B	C	D	E	4	5	6	7	8	9	0	1	2	3	16
2	3	4	5	6	A	B	C	D	E	12	13	14	15	16	0	1	8	9	10	11	7
11	12	13	14	A	B	C	D	E	3	4	5	6	7	8	9	10	16	0	1	2	15
3	4	5	A	B	C	D	E	11	12	13	14	15	16	0	1	2	7	8	9	10	6
12	13	A	B	C	D	E	2	3	4	5	6	7	8	9	10	11	15	16	0	1	14
4	A	B	C	D	E	10	11	12	13	14	15	16	0	1	2	3	6	7	8	9	5
A	B	C	D	E	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	0	13
B	C	D	E	9	10	11	12	13	14	15	16	0	1	2	3	A	5	6	7	8	4
C	D	E	0	1	2	3	4	5	6	7	8	9	10	11	A	B	13	14	15	16	12
D	E	8	9	10	11	12	13	14	15	16	0	1	2	A	B	C	4	5	6	7	3
E	16	0	1	2	3	4	5	6	7	8	9	10	A	B	C	D	12	13	14	15	11
7	8	9	10	11	12	13	14	15	16	0	1	A	B	C	D	E	3	4	5	6	2
16	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	15	11	12	13	14	10
8	9	10	11	12	13	14	15	16	0	A	B	C	D	E	6	7	2	3	4	5	1
5	14	6	15	7	16	8	0	9	1	10	2	11	3	12	4	13	A	B	C	D	E
14	6	15	7	16	8	0	9	1	10	2	11	3	12	4	13	5	E	A	B	C	D
6	15	7	16	8	0	9	1	10	2	11	3	12	4	13	5	14	D	E	A	B	C
15	7	16	8	0	9	1	10	2	11	3	12	4	13	5	14	6	C	D	E	A	B
13	5	14	6	15	7	16	8	0	9	1	10	2	11	3	12	4	B	C	D	E	A

XIV. Computer Construction of O(10, t) Sets

In about fifteen years the effectiveness of computers in searching for orthogonal sets of latin squares of order ten has increased strikingly. Still the problem is so large that there seems to be little reason for optimism that the order ten problem can be completed by computers. More precisely, if (as most conversant with the problem consider quite plausible) no orthogonal triple of orthogonal latin squares of order ten exists, then the number of cases to consider seems too large for an exhaustive proof by computer to be achievable. The number of latin squares of order ten is astronomical.

About 1953 Paige and Tompkins [195=] programmed SWAC to search for squares orthogonal to a fixed latin square of order ten. A few hours of running produced no orthogonal square, and was regarded as a bit of experimental evidence for the truth of Euler's conjecture. Calculations based on the progress made in the search led to the extrapolation that over fifty million years of computer time would be required to search for all squares orthogonal to a latin square of order ten put into SWAC intially. (At about the same time a similar program was written and similar results obtained with MANIAC at Los Alamos; this attempt has not been reported in print.)

In 1959, after Euler's conjecture had been disproved for all orders $4t + 2 > 6$, Parker programmed UNIVAC 1206 to search for squares orthogonal to a latin square of order ten. The running time was sharply less than for SWAC or MANIAC, about thirty minutes for the majority of latin squares. This was accomplished by generating and storing all transversals of the input latin square, then searching for all ways to form latin squares from the list of transversals. (A transversal, or directrix, is a set of cells of a latin square, one in each row, one in each column, and one containing each digit.) The striking gain in speed over the earlier efforts occurred largely because the number of transversals of a typical latin square of order ten is roughly 850, much less than 10!; and, of course, the search was several levels deep. (SWAC and MANIAC were programmed to build up starts of latin squares toward orthogonal mates by filling in cells to form rows.)

There were two main outcomes from considerable running of Parker's 1206 program: 1) Orthogonal triples of order ten latin squares are not numerous; more precisely, only a small fraction, if any, order ten squares extend to triples. Some 400 latin squares were run. Some were random, some were computer output fed back as input and hence known to have an orthogonal mate, and some were considered interesting candidates for intuitive reasons by Parker and others. Not once did an exhaustive search for orthogonal mates of an input latin square include a pair orthogonal to one another. Mild evidence may be claimed supporting the opinion that no order-ten orthogonal triple exists. 2) Of a computer-generated sample of 100 random latin squares of order ten (program by R. T. Ostrowski), 62 have orthogonal mates. Thus, unlike triples, order ten orthogonal pairs are quite common. Euler's intuition for order ten was not only wrong, but in this sense wrong by a large margin. It was this finding which tempted Parker for a time to believe that repeated runs of the program should have a good chance of producing at least a triple, but many failures dimmed this optimism.

In 1967 John W. Brown programmed IBM 7094 to decide whether an input latin square of order ten can be extended to an orthogonal triple. The running time was one half minute. Almost needless to say, transversals again were generated. Searching for patterns of transversals toward extension to a triple produced a speed gain over the previous program for orthogonal pairs. Brown endeavored to get every drop of speed from the machine. As before, hundreds of input order-ten latin squares produced no orthogonal triple.

XV. On the Equivalence of O(n,t) Sets With Other Combinatorial Systems

XV.0. Summary

In this section we have densely summarized some of the results obtained by author and at least fourteen others in order to demonstrate the importance of the theory of mutually orthogonal latin squares. We have shown that fourteen well-known and important combinatorial systems with certain parameters are actually equivalent to a set of mutually orthogonal latin squares. A schematic representation of these equivalences has been demonstrated in four wheels which we have called "Fundamental Wheels of Combinatorial Mathematics".

XV.1. Introduction

The theory of mutually orthogonal latin squares owes its importance to the fact that many well-known combinatorial systems are actually equivalent to a set of mutually orthogonal latin squares; viz., finite projective plane, finite Euclidean plane, net, BIB, PBIB, orthogonal arrays, a set of mutually orthogonal matrices, error correcting codes, strongly regular graphs, complete graphs, a balanced set of ℓ-restrictional lattice designs, difference sets, Hadamard matrices, and an arrangement of non attacking rooks on hyperdimensional chess board. These combinatorial systems are unquestionably potent and effective in all branches of combinatorial mathematics, and in particular, in the construction of experimental designs. Therefore, a statement that the theory of mutually orthogonal latin squares is perhaps the most important theory in the field of experiment designs is not in the least exaggerated as far as this author is concerned.

Our purpose in this section is to demonstrate the relation of a set of mutually orthogonal latin squares with the above mentioned combinatorial systems. We shall present the essence of the known results available only in scattered literature in one theorem which we consider to be a "fundamental theorem of combinatorial mathematics". For the definitions of these combinatorial systems and the proof of the forthcoming theorem see the list of references given at the end of this paper.

XV. 2. Notation

For the sake of conciseness we introduce the following notations:

0) $O(n,t)$ denotes a set of t mutually orthogonal latin squares of order n.

1) $MOM(n,t)$ denotes a set of t mutually orthogonal $n \times n$ matrices.

2) $OA(n,t)$ denotes a set of orthogonal arrays of size n^2, depth t, n levels, and strength 2.

3) $Net(n,t)$ denotes a net of order n and degree t.

4) $Code(n,r,t;m)$ denotes a set of n code words each of length r such that any two code words are at least at Hamming distance $\geq t$ on an m-set Σ with m distinct elements. We remind the reader that such a code is also called $(t-1)$-error detecting code or $(t-1)/2$-error correcting code because such a code is capable of detecting up to $t-1$ errors and correct up to $(t-1)/2$ errors in each transmitted code word.

5) $PBIB(b,v,r,k,\lambda_1,\lambda_2)$ denotes a partially balanced incomplete block design with b blocks each of size k, v treatments with r replication of each, and association indices λ_1 and λ_2.

6) SR-Graph (A) denotes the strongly regular graph with incidence matrix A .

7) Non #(n,t) denotes an arrangement of n mutually non attacking rooks on the t-dimensional $n \times n$ chess board.

8) PG(2,s) denotes a finite projective plane of order s (not necessarily Desarguessian).

9) $\mathcal{E}(2,s)$ denotes a finite Euclidean plane of order s .

10) BIB(b,v,r,k,λ) denotes a balanced incomplete block design with b blocks each of size k, v treatments with r replications of each, and association index λ .

11) K-Graph (A) denotes the complete graph with incidence matrix A .

12) DIF(v,k,λ) denotes a difference set with parameters v, k, and λ .

13) BLRL(s) denotes a balanced set of ℓ-restrictional lattice design for s treatments. Note that a 1-restrictional balanced lattice designs for simply a BIB design.

14) HAD(n) denotes a symmetric normalized Hadamard matrix of order n .

Hereafter we also adopt the following two notations:

i) A $\Longleftrightarrow$ B means A implies B and B implies A .

ii) A $\Longrightarrow$ B means A implies B . Whether or not B implies A is undecided.

2. The Result:

Theorem

(a) For any pair of positive integers n and t we have:

1) $O(n,t) \iff MOM(n,t+2)$

2) $O(n,t) \iff OA(n,t+2)$

3) $O(n,t) \iff Net(n,t+2)$

4) $O(n,t) \iff Code(n^2,t+2,\ t+1;n)$

5) $O(n,t) \iff PBIB(n^2,n(t+2),n,t+2,O,1)$

6) $O(n,t) \iff$ SR-Graph (A) where A is the incidence matrix associated with PBIB in 5).

7) $O(n,t) \iff Non\ \#(n^2,n^{t+2})$.

(b) If $t = n-1$ then also:

8) $O(n,n-1) \iff PG(2,n)$

9) $O(n,n-1) \iff \mathcal{E}(2,n)$

10) $O(n,n-1) \iff BIB(n^2+n+1,\ n^2+n+1,\ n+1,\ n+1,1)$

11) $O(n,n-1) \iff Code(n^2+n+1,n^2+n+1,\ 2n;2)$

12) $O(n,n-1) \iff$ K-Graph (A) where A is the incidence matrix associated with BIB in 10)

13) $O(n,n-1) \iff DIF(n^2+n+1,\ n+1,1)$.

(c) If $n = p^m$ where p is a prime and m is a positive integer then also the following:

14) $O(p^m,p^m-1) \iff BLRL(p^m)$.

(d) <u>If</u> $n = 2r$ <u>and</u> $t = r-2$, $r \geq 3$ <u>then the following are also true:</u>

15) $O(2r, r-2) \Longrightarrow HAD(4r^2)$

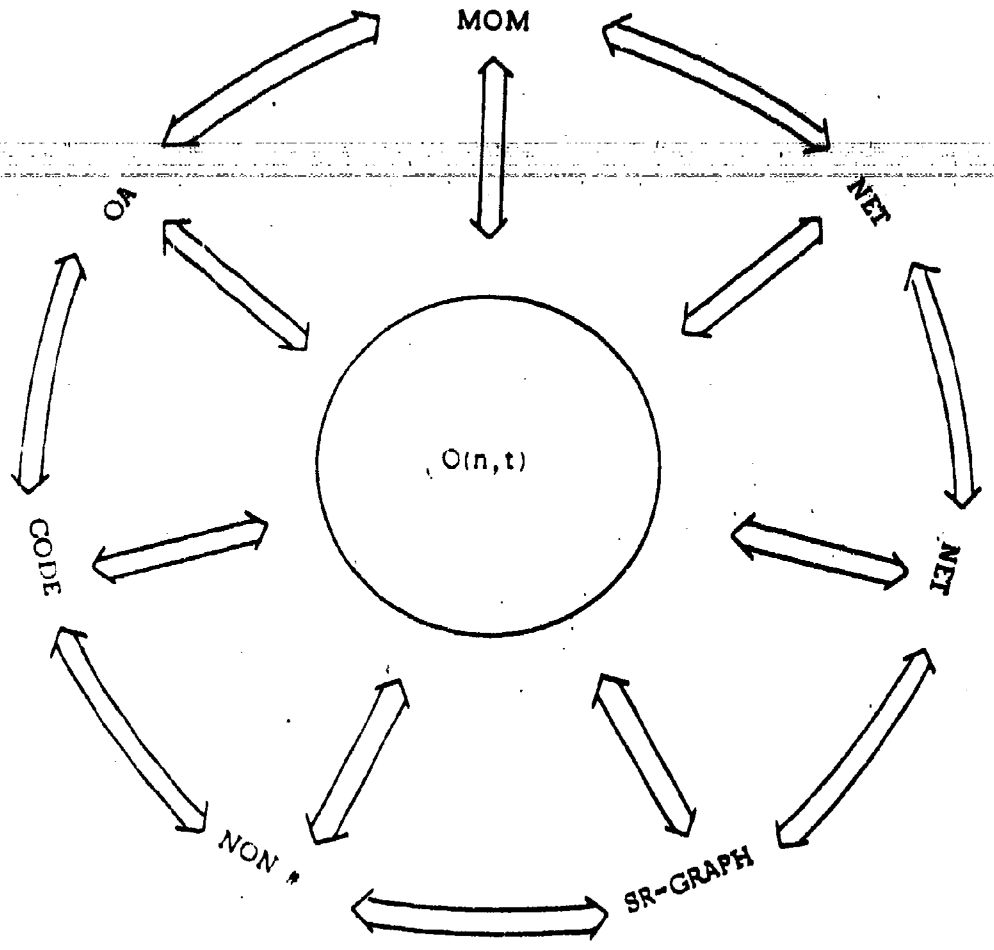
16) $O(2r, r-2) \Longrightarrow BIB(4r^2-1,\ 4r^2-1,\ 2r^2-1,\ 2r^2-1,\ r^2-1)$

17) $O(2r, r-2) \Longrightarrow Code(4r^2-1,\ 4r^2-1,\ 2r^2;2)$

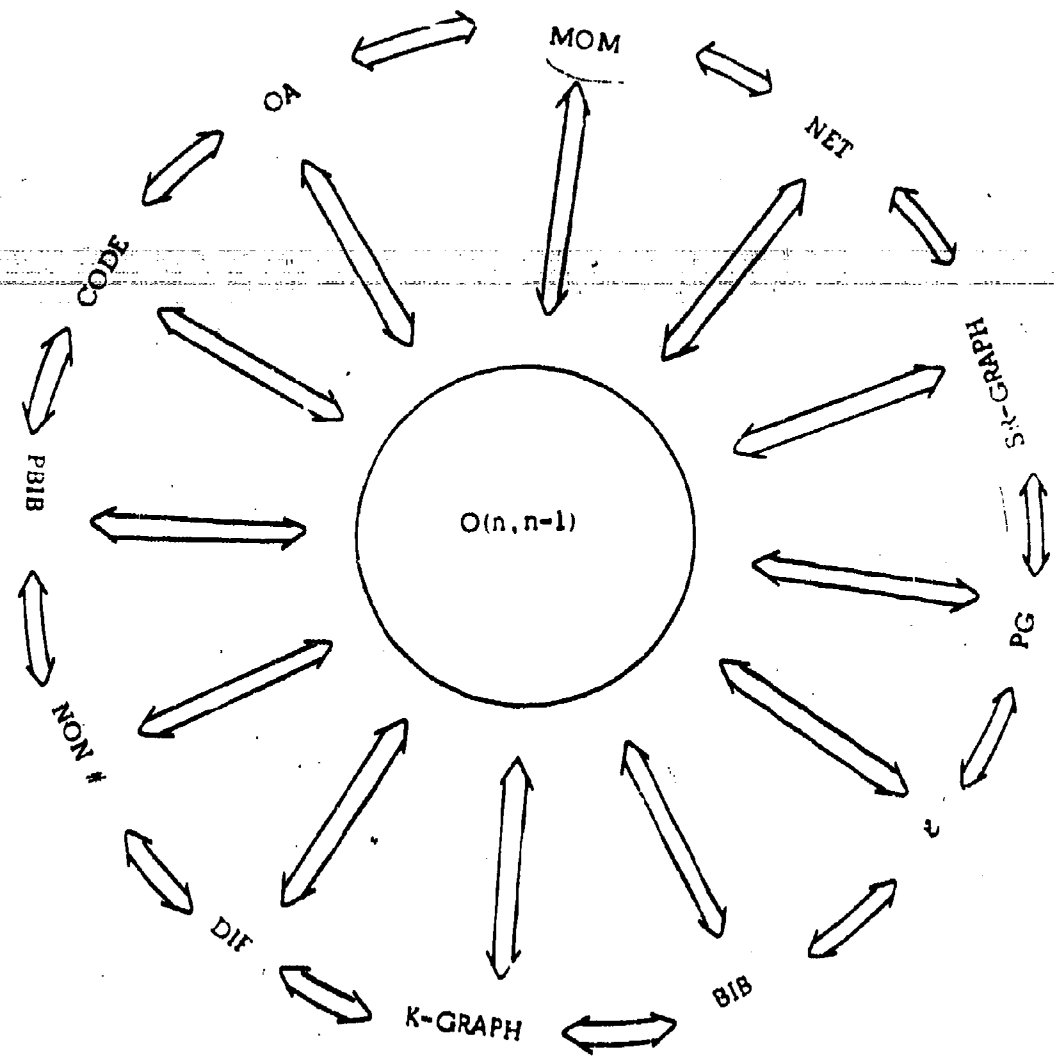
18) $O(2r, r-2) \Longrightarrow Code(8r^2,\ 4r^2,\ 2r^2;\ 2)$

19) $O(2r, r-2) \Longrightarrow DIF(4r^2-1;\ 2r^2-1,\ r^2-1)$.

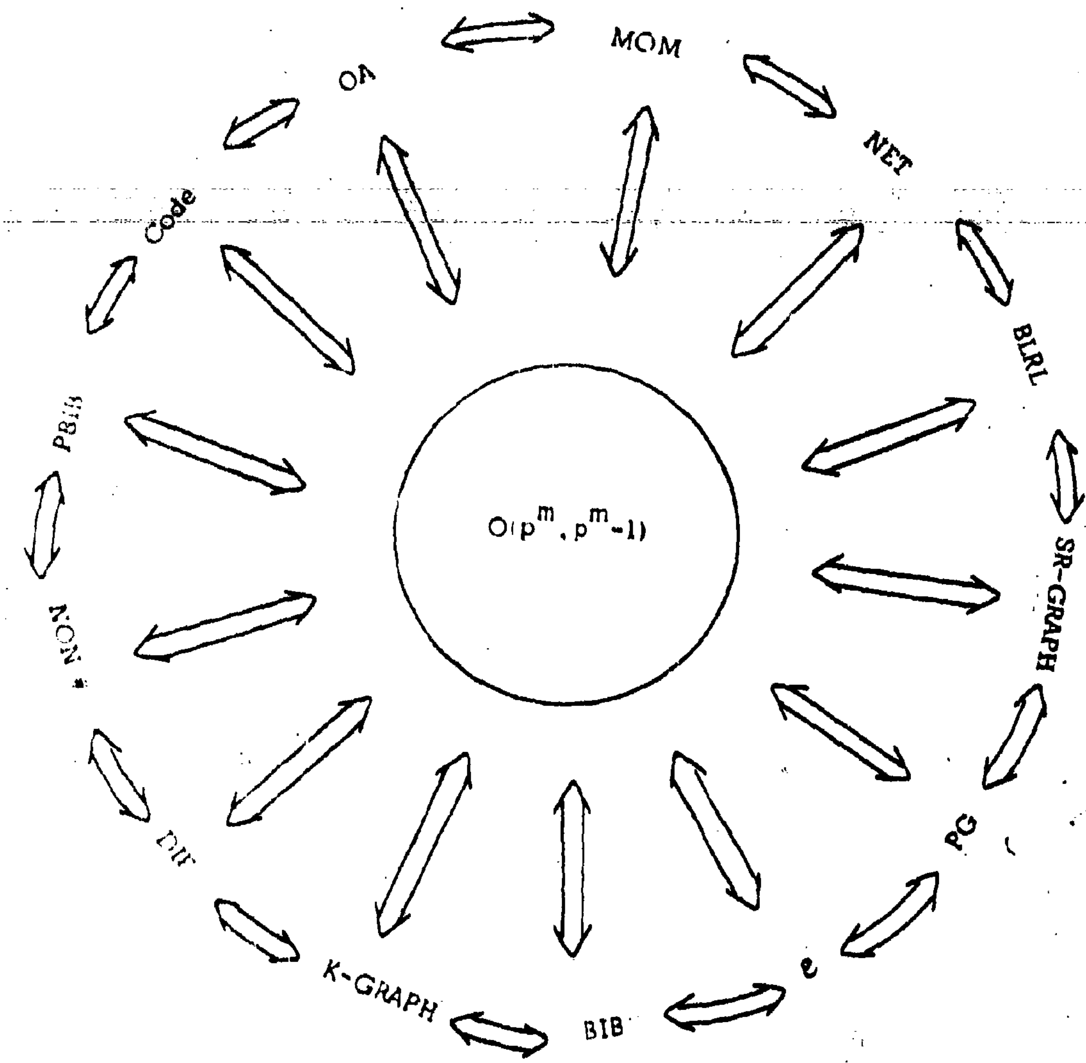
A complete schematic representation of this theorem can be demonstrated in four wheels which will be called "fundamental wheels of combinatorial mathematics". For the sake of compactness we shall omit the associated parameters with each system in these wheels except for $O(n,t)$. By knowing the values of n and t in the given $O(n,t)$ sets, then the reader can easily find the associated parameters with other systems in the wheels from the proper part of above theorem.



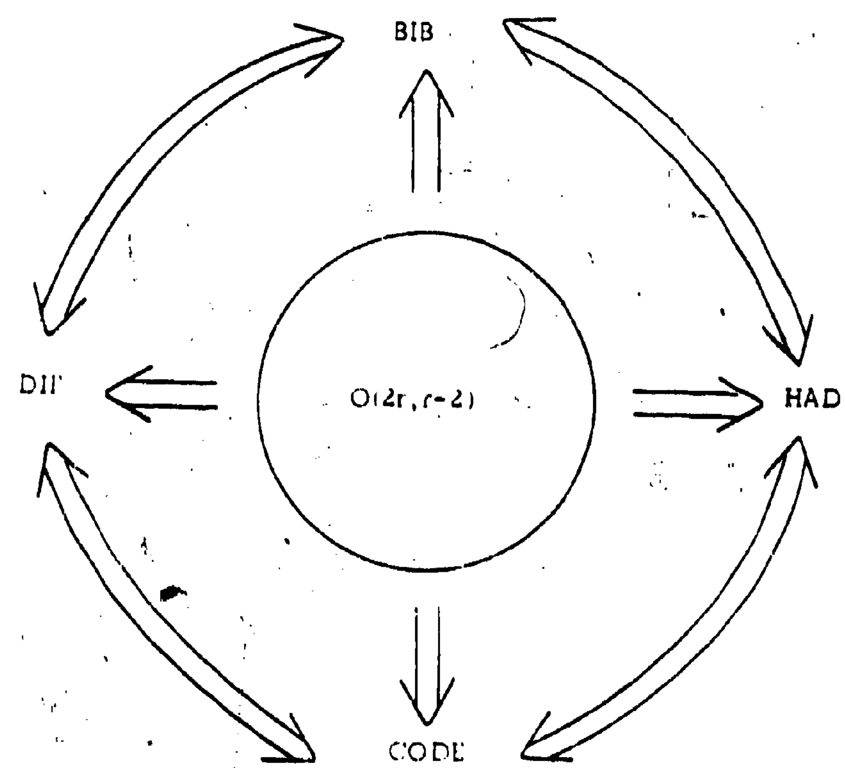
Wheel 1. For any positive integer n and t.



Wheel 2. For any positive integer n .



Wheel 3. For any prime p and positive integer m.



Wheel 4. For any positive integer $r \geq 3$.
(see also wheel 1)

XVI. Acknowledgements

This work was partially supported under the following research grants or contracts:

(i) Public Health Research Grant GM-05900, Cornell University

(ii) United States Army Contract Number DA-31-124-ARO-D-462, University of Wisconsin

(iii) National Science Foundation Grant GP-12881, Michigan State University

(iv) Office of Naval Research Contract N0014-A67-0305-0008, University of Illinois.

XVII. Literature Cited

1. Beckenbach, E. F. (ed.) [1964], Applied Combinatorial Mathematics, John Wiley and Sons, Inc., New York, London, Sydney.

2. Bose, R. C. [1939], On the application of the properties of Galois Fields to the problem of construction of Hyper-Graeco-Latin squares, Sankhya 3: 323-338.

3. Bose, R. C. [1939], On the construction of balanced incomplete block designs, Annals of Eugenics 9: 343-399.

4. Bose, R. C. [1950], A note on orthogonal arrays (abstract), Annals of Mathematical Statistics 21: 304-305.

5. Bose, R. C. [1963], Strongly regular graphs, partial geometrics and partially balanced designs, Pacific Journal of Mathematics 13: 389-419.

6. Bose, R. C. Chakravarti, I. M., and Knuth, D. E. [1960], On methods of constructing sets of mutually orthogonal Latin squares using a computer, I. Technometrics 2: 507-516.

7. Bose, R. C. and Clatworthy, W. H. [1955], Some classes of partially balanced designs, Annals of Mathematical Statistics 26: 212-232.

8. Bose, R. C. and Shrikhande, S. S. [1959], On the falsity of Euler's conjecture about the non-existence of two orthogonal Latin squares of order 4t + 2, Proceedings of the National Academy of Sciences, U.S.A. 45: 734-737.

9. Bose, R. C. and Shrikhande, S. S. [1960], On the construction of pairwise orthogonal Latin squares and the falsity of a conjecture of Euler, Transactions of the American Mathematical Society 95: 191-209.

10. Bose, R. C., Shrikhande, S. S. and Parker, E. T. [1960], Further results on the construction of mutually orthogonal Latin squares and the falsity of Euler's conjecture, Canadian Journal of Mathematics 12: 189-203.

11. Bruck, R. H. [1951], Finite nets, I. Numerical invariants, Canadian Journal of Mathematics 3: 94-107.

12. Bruck, R. H. [1963], Finite nets, II. Uniqueness and embedding, Pacific Journal of Mathematics 13: 421-457.

13. Bruck, R. H. [1963], What is a loop? in "Studies in Modern Algebra" (editor, A. A. Albert), Math. Assoc. Amer. and Prentice Hall, Englewood Cliffs, New Jersey, pp. 59-99.

14. Dembowski, P. [1968], Finite Geometries, Springer-Verlag, Berlin, Heidelberg, New York.

15. Euler, L. [1782], Recherches sur une nouvelle espèce des quarrés magiques, Verh. Zeeuwsch Genoot. Wetenschappen, Vlissigen 9: 85-239.

16. Fisher, R. A. and Yates, F. [1957], Statistical Tables for Biological, Agricultural and Medical Research, 5th edition (first edition in 1938), Oliver and Boyd, Edinburgh and London.

17. Golomb, S. W. and Posner, E. C. [1964], Rook domains, Latin squares, affine planes, and error-distributing codes, IEEE Transactions on Information Theory IT-10: 196-208.

18. Hall, M. [1943], Projective planes, Transactions of the American Mathematical Society 54: 229-277.

19. Hall, M. Jr. [1967], Combinatorial Theory, Blaisdell Publishing Co., Waltham, Massachusetts, Toronto and London.

20. Hall, M. and Paige, L. J. [1955], Complete mappings of finite groups, Pacific Journal of Mathematics 5: 541-549.

21. Hedayat, A. [1968], On Singer 1-permutation, Unpublished paper, Biometrics Unit, Cornell University, Ithaca, New York.

22. Hedayat, A. [1969], On the theory of the existence, non-existence, and the construction of mutually orthogonal F-squares and latin squares, Ph.D. Thesis, Cornell University, Ithaca, New York.

23. Hedayat, A. [1969], On the equivalence of a set of mutually orthogonal latin squares with other combinatorial systems, Paper RM-237, Dept. of Statistics and Probability, Michigan State University, East Lansing, Michigan.

24. Hedayat, A. and Federer, W. T. [1969], An application of group theory to the existence and non-existence of orthogonal Latin squares, Biometrika 56(3):

25. Hedayat, A. and Federer, W. T. [1969], On the equivalence of Mann's group automorphism method of constructing an O(n, n-1) set and Raktoe's collineation method of constructing a balanced set of ℓ-restrictional prime-powered lattice designs, (unpublished paper), Biometrics Unit, Cornell University, Ithaca, New York.

26. Hedayat, A. and Seiden, E. [1969], On a method of sum composition of orthogonal latin squares. Paper No. RM-238, Dept. of Statistics and Probability, Michigan State University, East Lansing, Michigan.

27. Hedayat, A. And Seiden, E. [1969], Some contributions to the theory of F-squares, Unpublished paper, Cornell University and Michigan State University.

28. Johnson, D. M., Dulmadge, A. L., and Mendelsohn, N. S. [1961], Orthomorphisms of groups and orthogonal Latin Squares I., Canadian Journal of Mathematics 13: 356-372.

29. Kempthorne, O. [1952], The Design and Analysis of Experiments, John Wiley and Sons, Inc., New York.

30. Levi, F. W. [1942], Finite Geometrical Systems, University of Calcutta, Calcutta, India.

31. MacNeish, H. F. [1922], Euler's squares, Annals of Mathematics 23: 221-227.

32. Mann, H. B. [1942], The construction of orthogonal Latin squares, Annals of Mathematical Statistics 13: 418-423.

33. Mann, H. B. [1943], On the construction of sets of orthogonal Latin squares. The Annals of Mathematical Statistics 14: 401-414.

34. Mann, H. B. [1944], On orthogonal Latin squares, Bulletin of the American Mathematical Society 50: 249-257.

35. Mann, H. B. (ed.) [1968], Error Correcting Codes, John Wiley and Sons, Inc., New York, London, Sydney, Toronto, pp. xii + 231.

36. Paige, L. J. [1951], Complete mappings of finite groups, Pacific Journal of Mathematics 1: 111-116.

37. Paige, L. J. and Tompkins, C. B. [1960], The size of the 10 × 10 orthogonal latin square problem, Proceedings of the Tenth Symposium in Applied Mathematics of the American Mathematics Society 10: 71-83.

38. Parker, E. T. [1959], Construction of some sets of mutually orthogonal latin squares, Proceedings of the American Mathematics Society 10: 946-949.

39. Parker, E. T. [1962], On orthogonal latin squares, Proceedings of Symposia in Pure Mathematics, American Mathematics Society 6: 43-46.

40. Peterson, W. W. [1961], Error-Correcting Codes, John Wiley and Sons, Inc., New York.

41. Raktoe, B. L. [1967], Application of cyclic collineations to the construction of balanced l-restrictional prime powered lattice designs, Annals of Mathematical Statistics 38: 1127-1141.

42. Raktoe, B. L. [1969], Combining elements from finite fields in mixed factorials, The Annals of Mathematical Statistics 40: 498-504.

43. Raktoe, B. L. and Federer, W. T. [1969], On irregular fractions of an s^m factorial, Unpublished paper, Biometrics Unit, Cornell University, Ithaca, New York.

44. Ryser, H. J. [1963], Combinatorial Mathematics, The Carus Mathematical Monographs No. 14, Mathematical Association of America and John Wiley and Sons, Inc., New York.

45. Sade, A. [1958], Groupes des orthogonaux, Publicationes Mathematicae, 5: 229-240.

46. Silverman, R. [1960], A metrization for power sets with applications to combinatorial analysis, Canadian Journal of Mathematics 12: 158-176.

47. Singer, J. [1960], A class of groups associated with Latin squares, The American Mathematical Monthly 67: 235-240.

48. Stevens, W. L. [1939], The completely orthogonalized Latin squares, Annals of Eugenics 9: 83-93.

49. Tarry, G. [1899], Sur le probleme d'Euler des n^2 officiers, L'Intermédiaire des Mathématiciens 6: 251-252.

ON CONFIDENCE LIMITS FOR THE PERFORMANCE OF A SYSTEM WHEN FEW FAILURES ARE ENCOUNTERED

Sam C. Saunders
Boeing Scientific Research Laboratories
Seattle, Washington

SUMMARY. In some situations encountered today components, or assemblies, are so reliable that no failures are observed within the time available for testing. This can pose a problem in both interpretation and analysis. We consider here the problem of determining lower confidence bounds on the reliability of a complex system, such as the Saturn 1-C, when each component is assumed to have an exponential life and different components have different multiplicities within the system. We discuss the assumptions necessary to obtain confidence limits using the likelihood of the data when only a few failures are encountered. The bounds resulting from two models are compared. The first model is Bayesian with uniform prior distribution of the failure rates. The second model regards the failure rates virtually as unknown constants. Here the argument is made that models of the first type are deficient in several regards in comparison with the second.

0. INTRODUCTION. The problem of determining the probability of successful operation of a large complex system when one has data only on the reliability of the components has, over the past decade, been the subject of many investigations. However, much of the literature was of a proprietary nature and was never published, for example, see [1], [2] and the references there.

Some of the studies, see [7] and [9], were based on an asymptotic theory for which the precision of the approximation is unknown. Currently, much of the analysis is based on Bayesian methods utilizing subjective prior assumptions, see [12] and [13].

Because estimating the probability of failure under some models requires that at least one failure be observed, the statistician may be placed in the uncomfortable position of having less confidence in his estimates of reliability when fewer failures obtain. Ultimately, when the system becomes near perfect and no failures are observed the statistician has no confidence if his procedures are necessarily based on failure analysis. In this unsatisfactory situation it is an understandable reaction of persons with good engineering judgment and statistical intuition, to form a distrust of statistical inference and its "numerologists," see [3]. Some recent surveys have been made to determine the most useful and applicable procedures for current needs. One of the most comprehensive is [8].

In this note we examine some statistical techniques which do not depend upon the sampling method yet are applicable when there is a paucity of observed failures among the components which have been tested. The archtypical situation for this study will be the Saturn 1-C and the data which was available prior to the first launch, as given in Table I.

The remainder of this article has been reproduced photographically from the author's manuscript.

1. THE BASIC MODEL

Consider a complex system designed to perform in a specified manner when all of its components, grouped in assemblies, are operating. At question is the confidence we can have that it will perform adequately for a specified time.

We shall consider the reliability of the system for a specified time t as being given by

$$h(\underset{\sim}{\lambda},t) = \exp\left\{-\sum_{i=1}^{m} \lambda_i \omega_i t\right\} \tag{1.1}$$

where the ω_i are known weighting factors for the unknown parameters $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_m)$. The problem is obtaining a lower confidence bound on (1.1) from the limited amount of data about the λ_i for $i=1,\ldots,m$.

The particular form (1.1) can arise in several ways. The first we mention is that it is itself a lower bound on system reliability.

If the time until failure of the i^{th} component is exponential with unknown hazard rate λ_i and these m components are in a coherent system, then there exists a set of integers $q_1,\ldots,q_m$, where q_i is the multiplicity of the i^{th} component within the system of order $\sum_1^m q_i$, which can be used to obtain a lower bound on the reliability of the system at any time $t > 0$. This lower bound is $h(\underset{\sim}{\lambda},t)$ as given in (1.1) with ω_i replaced by q_i.

For a proof of this result, see [4].

An assembly containing several components can malfunction by having different components fail, e.g., a pressure system can either rupture or leak. These separate ways of failure are called "failure modes" in the current terminology, however, they correspond to malfunctions within the subsystem and need not necessarily cause a system failure.

We assume

1° The time until malfunction in each mode of a given assembly has constant hazard rate and all are independent.

Suppose we separate the possible modes of malfunction for each assembly into mutually exclusive, functionally independent, classes labeling the time until malfunction T_{ij} for the i^{th} mode of the j^{th} assembly. The time until malfunction of the assembly by any mode is $T_j = \min_i(T_{ij})$ and the hazard rate of T_j is $\lambda_j = \sum_i \lambda_{ij}$, with the obvious interpretation of λ_{ij} as the hazard rate of T_{ij}. Unfortunately, the system may have different vulnerability to such malfunctions depending upon the period within the mission phase.

Thus the second situation in which the form (1.1) can arise is a series system with such malfunctioning assemblies. Assume

2° Given a malfunction in the j^{th} mode of the i^{th} assembly during the k^{th} time interval $(t^{(k-1)}, t^{(k)})$ of a mission, system failure will result with known conditional probability β_{ijk}.

For a mission of fixed length t we define the beta factor for the i^{th} assembly, which is a known constant, by the equation

$$\beta_i = \sum_{j,k} \beta_{ijk}[t^{(k)} - t^{(k-1)}]$$

where $t^{(0)} = 0$, $\lim t^{(k)} = t$. Thus from 1°, 2° the probability of system safety over the time interval $(0,t)$ is $h(\lambda,t)$ as given in (1.1) with $\omega_i t$ replaced by β_i. Without loss of generality we shall henceforth assume that all time intervals are expressed in fractions of this fixed mission length t, to wit, assume $t = 1$.

Needless to say, in the practical example given, all of these complications occur simultaneously. Moreover, the determination of the q_i through minimal state reliability analysis is itself a non-trivial task, not to speak of the analysis of the malfunction modes and their effect upon mission success (or vehicle safety). All of this is a necessary prelude for the determiuation of the beta factors. But we shall assume this work has been completed so that $\omega_i = q_i \beta_i$ is known, consequently equation (1.1) holds with $t = 1$.

The problem that we wish to discuss is of another genre, namely the methods for utilisation of the data so as to determine a lower confidence bound for the reliability after the ω_i have been obtained. Since it is this aspect which is important we shall assume that $\beta_i \equiv 1$ for our data so as to further emphasize the differences between the two models under discussion. Of course this fictitiously makes the reliability estimate low.

2. THE DATA

In many cases the first (and sometimes the only) data one has concerning the reliability of the components comes from environmental tests. This test data must be reduced by engineering evaluation into the equivalent operational time during the given mission phase. These transformation factors, called K-factors, are used in industrial practice, see [1], to accelerate the testing and/or reduce its expense. Specifically, during the first phase of a mission a component may experience several types of vibration as well as several temperature and humidity changes. Consequently, testing the components in these separate environments must yield results requiring a transformation into the appropriate mission phase equivalent time. (The dangers of such a procedure are apparent but are taken in view of the exhorbitant cost of the alternative.)

We do not discuss this further but we merely point out that in such instances the data on the operational behavior of the components are not God-given, but rather are the construct of engineering knowledge and judgment.

Thus, the statistician is ultimately provided with data on all components in the form

$$x_i = (t_i, n_i) \qquad i=1,\ldots,m \tag{2.1}$$

where t_i is the total time, expressed in equivalent fractions of the mission length that the i^{th} component has been operated, and n_i is the total number of malfunctions of the i^{th} component during time t_i.

In the life qualification of components, it is usually the case that testing proceeds as long as there are funds available and this is usually neither until a fixed number of hours nor a fixed number of failures occur. Moreover, there are always extraneous circumstances which might terminate the testing program at any time. In view of this indeterminacy in the sampling procedure the treatment of the data that we adopt should not depend heavily upon a particular sampling scheme which might not obtain.

Assume that a number of identical components, say m, are put on test. What is observed at each trial is the random length of life, call it Y, when the component fails or the random time, say Z, at which the test is terminated for any reason other than failure of the component.

Thus we observe the event

$$[Y = y] \cap [Y < Z] \quad \text{or} \quad [Z = z] \cap [Y \geq Z]$$

which we note is not the minimum of Y and Z since we know whether Y or Z is observed. It is well known that the likelihood is independent of the sampling method for type I or type II censoring, i.e. stopping at either a fixed number of failures or after a fixed time".

We make the

<u>Remark</u>: If $(Z_1, \ldots, Z_m)$ is a vector of non-negative random variables, independent of all Y_i such that $[Y_i > Z_i]$, where $(Y_1, \ldots, Y_m)$ are themselves independently and identically distributed with common density function f and distribution F, then the likelihood of the event

$$\bigcap_{i=1}^{n} [Y_i \le Z_i][Y_i = y_i] \bigcap_{i=n+1}^{m} [Y_i > Z_i][Z_i = z_i],$$

where n ($\le m$) is the random number of failures observed, is of the form

$$C \prod_{i=1}^{n} f(y_i) \prod_{j=n+1}^{m} [1-F(z_j)] \tag{2.2}$$

and the constant C depends upon the outcomes $(y_1,\ldots,y_n)$ but not upon their distribution.

The proof is immediate. Let g be conditional density of $(Z_1,\ldots,Z_m)$ given $(Y_1,\ldots,Y_n)$ assuming $Z_1,\ldots,Z_m$ independent of $Y_{n+1},\ldots,Y_m$ on $\bigcap_{i=n+1}^{m} [Y_i > Z_i]$. Then the probability of the event specified above is

$$\int_{\{z_i>y_i,\ i=1,\ldots,n\}} \cdots \int \prod_{i=1}^{n} f(y_i) \prod_{j=n+1}^{m} [1-F(z_j)] g(z_1,\ldots,z_m|y_1,\ldots,y_n) dz_1,\ldots,dz_n$$

which upon simplification shows that

$$C = \int_{\{z_i>y_i,\ i=1,\ldots,n\}} \cdots \int g(z_1,\ldots,z_m|y_1,\ldots,y_n) dz_1,\ldots,dz_n$$

as claimed.||

Taking the data $x_i = (t_i,n_i)$ for the i^{th} of m assemblies, the i^{th} assembly having exponential life with hazard rate λ_i, and substituting into (2.2) we obtain the likelihood

$$p(x_i|\lambda_i) = C(x_i)\lambda_i^{n_i} e^{-\lambda_i t_i} \tag{2.3}$$

where $C(x_i)$ is independent of λ_i.

Because the likelihood is the same for this very general sampling situation and we feel that the data by it's nature requires such independence we favor methods of statistical analysis which depend upon the likelihood.

From Bayesian Principles, Lindley [8], pp. 1,2, the joint posteriori density of $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_m)$, based on the evidence $\underset{\sim}{x} = (x_1,\ldots,x_m)$ with likelihood given in (2.3), is

(2.4) $$f(\underset{\sim}{\lambda}|\underset{\sim}{x}) \propto \prod_{i=1}^{m} p(x_i|\lambda_i)\pi(\underset{\sim}{\lambda})$$

where π is the joint prior density of $\underset{\sim}{\lambda}$.

Two difficulties remain. One is to formulate a reasonable joint prior π and the second is then to calculate, other than symbolically, the posterior distribution of

(2.5) $$V = \Sigma\beta_i\lambda_i, \quad \text{say} \quad G(v|\underset{\sim}{x}) \quad \text{for} \quad v > 0.$$

Following the usual method, p. 15, Lindley, loc. cit., the value v_0, depending upon $\underset{\sim}{x}$ and ε, $0 < \varepsilon < 1$ such that $G(v_0|\underset{\sim}{x}) = \varepsilon$, provides a lower $100\varepsilon\%$ Bayesian confidence bound for the system reliability e^{-V} given $\underset{\sim}{x}$ of the form e^{-v_0} and

(2.6) $$P[\exp\{-\Sigma\beta_i\lambda_i\} > e^{-v_0}] = \varepsilon.$$

Essentially this method has been utilized to obtain confidence bounds on the reliability of certain systems and is presently the subject of much discussion, see [8]. In what follows we shall discuss two such methods and their reasonableness in dealing with the situation at hand.

3. A UNIFORM PRIOR

The first approach is to assume the special prior density

$$(3.1) \qquad \pi(\underset{\sim}{\lambda}) \equiv 1 \quad \text{for all} \quad \lambda_i > 0.$$

This assumption is justified by the so-called principle of insufficient reason: since we know nothing specific about π we have insufficient reason to take π anything but uniform. Strictly speaking π as defined in (3.1) is a non-probabilistic prior. But, of course, one could consider it proportional to an approximation to a prior density.

Substituting (3.1) into (2.4) we find

$$(3.2) \qquad f(\underset{\sim}{\lambda}|\underset{\sim}{x}) = \prod_{i=1}^{m} \frac{t_i(\lambda_i t_i)^{n_i} e^{-\lambda_i t_i}}{n_i!} \quad \text{for} \quad \lambda_i > 0.$$

The mathematical problem becomes that of finding the distribution of $V = \Sigma\beta_i\lambda_i$ where β_i are known constants and λ_i are gamma variates with known scale and shape parameters. To wit, each λ_j is $\Gamma(t_j, n_j+1)$ where $\Gamma(t,\nu)$ denotes the law with density, given $\nu > 0$

$$(3.3) \qquad \frac{1}{\Gamma(\nu)} t^{\nu} x^{\nu-1} e^{-tx} \quad \text{for} \quad x > 0.$$

We also quote two related results, see p. 46ff, Feller [7].

If λ_j is $\Gamma(t_j,\nu_i)$, then $t_j\lambda_j$ is $\Gamma(1,\nu_j)$.

If $\nu_j > 1$, then $\lambda_j = \lambda_j' + \lambda_j''$ in distribution where λ_j' is $\Gamma(t_j,1)$ independent of λ_j'' which is $\Gamma(t_j,\nu_j-1)$.

Thus by the first remark we see that in distribution $V = \sum_{j=1}^{m} b_j\lambda_j$

where $b_j = \beta_j/t_j = 1/\tau_j$, $j=1,\ldots,m$ and each λ_j is now $\Gamma(1,n_j+1)$. By the second remark, for the data given in Table I where we have at most two failures, we see that in distribution

$$(3.4) \qquad V = \sum_1^m b_i Z_i + \sum_1^r b_j Y_j + \sum_1^s b_k X_k$$

where s is the number of components with two failures during testing

$r-s$ is the number of components with one failure during testing

$m-r-s$ is the number of components with no failures during testing

and X_k, V_j, Z_i are all independent $\Gamma(1,1)$, i.e., exponential with unit mean, variates.

We now quote a result proved, for example, in [11] as a

<u>Lemma 1</u>: If $Z_1,\ldots,Z_k$ are independent exponential random variables with unit mean, then for $b_i > 0$, all distinct, we have

$$(3.5) \qquad P[\sum_1^k b_i Z_i > u] = \sum_{j=1}^k B_j^{(k)} e^{-u/b_j}$$

where $B_1^{(1)} = 1$ and for $k \geq 2$

$$(3.6) \qquad B_j^{(k)} = \prod_{\substack{i=1 \\ i\neq j}}^k \frac{b_j}{b_j - b_i} \quad \text{for} \quad j=1,\ldots,k.$$

Also these recursion relations hold

$$B_j^{(k)} = B_j^{(k-1)} b_j/(b_j - b_k), \quad j=1,\ldots,k-1 \quad \text{and} \quad B_k^{(k)} = 1 - \sum_{j=1}^{k-1} B_j^{(k)}.$$

Also we have

<u>Lemma 2</u>: The distribution of $\sum_1^m b_i Z_i + \sum_1^r b_j Y_j$ is

$$\sum_{i=1}^{m} \sum_{j=1}^{r} B_i^{(m)} B_j^{(r)} \{1-e^{-\tau_j t} - \tau_j \psi(i,j,t)\} \quad \text{for} \quad t > 0,$$

where $\tau_j = 1/b_j$ for $j=1,\ldots,\max(m,r)$

$$\psi(i,j,t) = \int_0^t \exp\{-\tau_i(t-y)-\tau_j y\} dy$$

$$= \begin{cases} te^{-\tau_i t} & \text{if} \quad i = j \\ \dfrac{e^{-\tau_i t}-e^{-\tau_j t}}{\tau_j - \tau_i} & \text{if} \quad i \neq j. \end{cases}$$

The proof is accomplished by the convolution of two distributions each of the form given in Lemma 1.

Consider the more general definition

$$V_k = \sum_{j=1}^{k} \sum_{i=1}^{m_j} b_i X_{ij} \tag{3.7}$$

where the X_{ij} are all independent exponential variates with unit mean. Let V_k have distribution F_k, then

$$V_k = V_{k-1} + \sum_{i=1}^{m_k} b_i X_{ik}.$$

Defining $\bar{F} = 1-F$ with any affixes, and taking τ_i as given in Lemma 2,

$$\bar{F}_k(u) = \int_0^\infty P\left[\sum_{i=1}^{m_k} b_i X_{ik} > u-v\right] dF_{k-1}(v)$$

$$= \bar{F}_{k-1}(u) + \sum_{i=1}^{m_k} B_i^{(m_k)} \left\{\int_0^u e^{-(u-v)\tau_i} dF_{k-1}(v)\right\}.$$

But the quantity in braces in the equation above becomes

$$\{\cdots\} = \bar{F}_{k-1}(u) - e^{-u\tau_i} - \tau_i \int_0^u \bar{F}_{k-1}(v) e^{-(u-v)\tau_i} dv.$$

Hence we have shown the following:

Lemma 3: The survival probability of V_k as defined in equation (3.7) is given in terms of the survival probability of V_{k-1} as

$$\bar{F}_k(u) = \sum_{i=1}^{m_k} B_i^{(m_k)} \left[e^{-u\tau_i} + \tau_i \int_0^u \bar{F}_{k-1}(v) e^{-(u-v)\tau_i} dv \right]. \tag{3.8}$$

Note that (3.8) can be used to prove Lemma 2 and used recursively to find the distribution of V_k for small k. Thus we now have a computationally feasible method for the calculation of the distribution of V, called G. Using the lemmas above a machine program was written for the IBM 360, using double precision for the computation of the $B_j^{(k)}$, which tabulates the distribution of V in the region of interest. Using the data presented in Table I, this distribution is graphed in Figure 1.

For example, we find that if $v_0 = 12.8$, $G(v_0) = .95$. Thus a lower 95% Bayesian confidence limit for the system reliability is $e^{-12.8} \doteq 10^{-5}$.

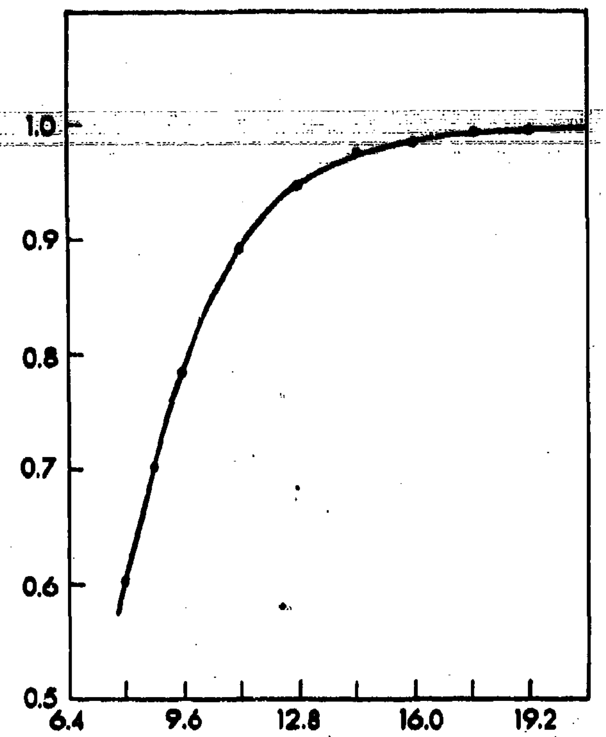


Figure 1

Graph of the distribution F of the random variable $V = \Sigma\beta_i\lambda_i$ in the region near unity. If v is the abscissa value, the value $F(v)$ of the ordinate is the confidence the system reliability exceeds e^{-v}.

4. A CRITICAL DISCUSSION

A word about the computation necessitated by this method. It is clear from Table I that the differences of the τ_j are neither small nor uniform. A glance at the formula for the $B_j^{(k)}$ in equation (3.6) shows that in absolute value they can become very large for such cases. (In fact, for such data as we have for 67 components, values as high as 10^{20} are not impossible.) Since all $B_j^{(k)}$ summed over j must add to unity, some must be positive and some negative. However, because of the nature of machine decimal arithmetic, the summands will be rounded off and the machine cumulate the error. We should, by definition, have $G(0) = 0$ but computationally we do not. For example, referring to Figure 1, the machine value at $v = 3.2$ was $G(3.2) = .699 \times 10^{-3}$ but at $v = 2.4$, $G(2.4) = -.123 \times 10^{-1}$ with wider fluctuations for smaller values of v. Fortunately, we are interested in those values of the argument for which $G(v)$ is near one and the values of v necessarily become large enough to eliminate the errors due to this circumstance.

However, this is merely a limitation due to the accuracy of the method of computation which was adopted. We feel there is a much more primary objection.

Without any real loss of clarity to the fundamental ideas, let us fix our attention on an assembly with two separate modes of malfunction with hazard rates λ_1 and λ_2, say. Suppose this assembly was operated for a time t and no failure of either type was observed. By using the uniform prior of (3.1) the posterior distribution of the component hazard rate $\lambda(= \lambda_1+\lambda_2)$ considering the component as a unit, is

(4.1) $P[\lambda < a] = P[e^{-\lambda} > e^{-a}] = 1-e^{-at}$ for $a > 0$

so that e^{-a} is a lower bound on the reliability for a mission of unit length and the confidence level is $P[\lambda < a]$.

But on the other hand, by considering the posterior distribution of $\lambda_1 + \lambda_2$, using data $x_1 = x_2 = (t,0)$ and the principle of insufficient reason to apply the uniform prior for each mode, we have

(4.2) $P[\lambda < a] = 1-e^{-at}-ate^{-at}$ for $a > 0$

as the posterior distribution of the hazard rate λ of the assembly. But notice that (4.2) is less than (4.1) which was the distribution from the same data for the same assembly.

The point we are making is simply this: For a series system, with no component failing during test time $t > 0$, the confidence in the reliability of the system should be the same as that for each component, since the system and the components both experience the same operational time t without failure.

Our criticism of the former method is that the confidence in the reliability does not depend only upon the data, it also depends upon the arbitrary designation of component or assembly. If we arrive at different answers when using the same data, then something must be wrong.

To continue this point further, let us suppose that we have a series system with separate malfunction modes with hazard rates $\lambda_1,\ldots,\lambda_m$ each of which has acquired the same operational experience, namely $x_i = (t, 0)$ for $i=1,\ldots,m$, i.e., no failures during operation for a length of time t. Again by using the uniform prior density we have the distribution of $\lambda = \lambda_1+\cdots+\lambda_m$ as

$$1 - \sum_{j=1}^{m} \frac{e^{-at}(at)^j}{j!} \tag{4.3}$$

which approaches zero as m approaches infinity regardless of the value of $ta > 0$.

Of course, (4.2) results from a different specification of the Bayesian model. The point is that each specification of another independent component will always result in a different posterior distribution. (Needless to say, a different prior will lead to a different posterior density for λ as well.)

Moreover, it is clear that almost any choice of prior density of λ which is the product of independent prior densities for each λ_i will result in a confidence level which is essentially the same as that given in (4.3), to wit so low as to be nonsensical for m large.

One modification suggested is to assume functional dependence with statistical independence, among the prior densities of λ_i. One such is to take the (conjugate) prior density of λ_i as $\Gamma(u_i,v_i)$ for $i=1,\ldots,m$, using here the notation of (3.5), subject to the constraint

$$v_i > 0, \qquad \sum_1^m v_i = 1. \tag{4.4}$$

Combining this with the likelihood of the form (2.3) shows

$$f(\lambda|x) \propto \prod_{i=1}^{m} [\lambda_i^{n_i+v_i-1} e^{-\lambda_i(u_i+t_i)}]$$

subject to (4.4) above. Thus the posterior density of λ_i is $\Gamma(u_i+t_i, n_i+v_i)$ and thus the mathematical problem becomes that of finding the distribution of $U = \Sigma c_i\lambda_i$ where $c_j = \beta_j/(u_j+t_j)$ and each λ_j is $\Gamma(1,n_j+v_j)$ for $i=1,\ldots,m$ subject to (4.4).

Because $n_j + v_j$ is not an integer we are faced with an analytic and computational problem beyond that of the preceding section. However, it is clear that this artiface does introduce enough degrees of freedom that proper choice of u_j, maintaining the restriction (4.4), can yield reliabilities of not unreasonable size. We do not pursue it further. The difficulty, making such an assumption untenable, is that ones prior knowledge about the reliability of a component should neither depend upon the prior densities of the other components in any way nor upon how many of them there are. These prior densities should be independent in every sense.

5. THE DEGENERATE PRIOR

The basis of Bayesian methodology is to regard the λ_i as random variables having a distribution which is to be constructed from prior knowledge. The more commonly accepted point of view is that the λ_i are unknown constants about which inference must be made. This is now the point of view which we essentially adopt.

If the λ_i were unknown positive real numbers, then there would exist a constant of proportionality between any two λ_i's which would be fixed, even though it was unknown.

Thus we make the assumption

3° There exists a constant of proportionality, say α_{ij}, between any two λ_i and λ_j.

If we have m different modes of malfunction, we define α_i for $i=1,\ldots,m$ as the *probability of malfunction in the* i^{th} *mode given that a malfunction in the system has occurred.* One sees that

$$\alpha_i = P[T_i < t \mid \sum_{j=1}^{m} [T_j < t]],$$

where we made the convention that the summation of events denotes the disjoint union. It follows that $\alpha_{ij} = \alpha_i/\alpha_j$ where

(5.0.1) $$\alpha_i = \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j} \qquad i=1,\ldots,m.$$

Thus 3° is equivalent to taking the prior distribution, say $\Pi(\underset{\sim}{\lambda})$, to be singular with all measure concentrated along a ray out from the origin with the direction of the ray determined by the constants of proportionality. Specifically, we assume

$$(5.1)\quad d\Pi(\lambda_1,\ldots,\lambda_m)\begin{cases} > 0 & \text{if some} \quad \lambda_i = \dfrac{\alpha_i\lambda_i}{\alpha_j} \quad \text{for} \quad i \neq j \\ = 0 & \text{otherwise.}\end{cases}$$

In the case $m = 2$, $\Pi(\lambda_1,\lambda_2)$ is zero everywhere but along the ray $\lambda_2 = \frac{\alpha_2}{\alpha_1}\lambda_1$ out from the origin in the (λ_1,λ_2) plane. We wish to find the posterior density of $\sum_1^m \beta_i\lambda_i$. We make the change of variables $\rho_i = \beta_i\lambda_i$ and by (2.3) and $\tau_i = t_i/\beta_i$ we have

$$f(\underset{\sim}{\rho}|x) \prec \sum_{i=1}^{m} (\tau_i\rho_i)^{n_i} e^{-\tau_i\rho_i} d\Pi^*(\rho_1,\ldots,\rho_m)$$

where

$$\Pi^*(\rho_1,\ldots,\rho_m) = \Pi(\frac{\rho_1}{\beta_1},\ldots,\frac{\rho_m}{\rho_m}).$$

Thus

$$(5.2)\quad d\Pi^*(\rho_1,\ldots,\rho_m)\begin{cases} > 0 & \text{if some} \quad \rho_i = \dfrac{\alpha_i\beta_i\rho_i}{\alpha_j\beta_j} \quad \text{for} \quad i \neq j \\ = 0 & \text{otherwise.}\end{cases}$$

The density we seek is proportional to

$$(5.3)\quad \int\underset{\{\underset{\sim}{\rho}:\Sigma\rho_i=a\}}{\cdots}\int \prod_{i=1}^{m} (\tau_i\rho_i)^{n_i} e^{-\tau_i\rho_i} d\Pi^*(\rho_1,\ldots,\rho_m).$$

Consider the line in m-space

$$\ell(\rho_1) = (\rho_1, \frac{\alpha_2\beta_2}{\alpha_1\beta_1}\rho_1,\ldots, \frac{\alpha_m\beta_m}{\alpha_1\beta_1}\rho_1).$$

By equation (5.2) all the mass of Π^* is concentrated along the ray $\ell(\rho_1)$ for $\rho_1 > 0$. In effect the only quantity that has a distribution is ρ_1 and we shall later see it makes no difference what this distribution is as long as it has support on $(0,\infty)$. This line intersects the plane $\sum_1^m \rho_i = a$ at a single point, namely ρ_1 such that

$$\rho_1 + \sum_{i=2}^{m} \frac{\alpha_i\beta_i}{\alpha_1\beta_1}\rho_1 = a$$

and solving for ρ_1 we find $\rho_1 = \gamma_1$ where we define

$$\gamma_i = \frac{\alpha_i\beta_i}{\sum_{j=1}^{m} \alpha_j\beta_j} \qquad i=1,\ldots,m. \tag{5.3.1}$$

Then the value of ρ_i at the point of intersection of the line ℓ with the plane $\Sigma\rho_i = a$ is $\rho_i = a\gamma_i$ for $i=1,\ldots,m$. Since all the measure of Π^* is concentrated along the line ℓ, the integration over the plane in (5.3) yields a single value at the singularity of the measure Π^*. It follows that the density we seek is proportional to the value of the integrand at that point, namely

$$\prod_{i=1}^{m} (\tau_i a\gamma_i)^{n_i} \exp\{-\sum_{i=1}^{m} \tau_i a\gamma_i\}.$$

If we define

$$\theta = \sum_{i=1}^{m} \tau_i\gamma_i = \frac{\Sigma t_i\alpha_i}{\Sigma\beta_i\alpha_i} \quad \text{and} \quad k = \sum_1^m n_i, \tag{5.4}$$

which are, respectively, a weighted mean of the t_i, and the total number of failures, we can write the posterior density of $V = \sum_1^m \rho_i$ as

$$(5.5) \qquad \frac{\theta(\theta a)^k e^{-a\theta}}{k!} \quad \text{for} \quad a > 0.$$

The distribution then is

$$(5.6) \qquad P[V < u] = \int_0^{\theta u} \frac{s^k e^{-s}}{k!}\, ds = 1 - \sum_{j=0}^{k} \frac{e^{-\theta u}(\theta u)^j}{j!}$$

which we recognize as a Chi-square distribution. If we set

$$(5.7) \qquad u = \frac{1}{2\theta}\, \chi_\varepsilon^2(2k+2),$$

where $\chi_\varepsilon^2(m)$ is the $100\varepsilon^{th}$ percentile of the Chi-square distribution with m degrees of freedom, we have e^{-u} providing a lower confidence bound of level ε.

We note that the computation for this method is trivial. We compute only the two quantities k and θ and then from a table of the Chi-square distribution calculate u and e^{-u}.

Unfortunately, equation (5.7) gives the confidence bound e^{-u} in terms of the alpha factors which are still unknown. Nonetheless, based upon the objective model that the failure rates are virtually unknown constants, we do arrive at (5.7) and knowledge concerning the α_i is what is needed to determine the confidence limit. However, this does not necessarily mean that the values of λ_i

need be known, for example, it is sufficient that their ratios be known. Perhaps in some instances engineering experience might be able to classify all the failure rates as multiples of fixed one, say the lowest, at least in a conservative manner.

Disregarding for the present the computation of θ, this method does obviate some of the conceptual difficulties which the preceding method possessed.

Firstly, the confidence bound is the same regardless of how the components are apportioned to subsystems within the system. In particular, if $\tau_1 = \cdots = \tau_m$, we obtain the same density of $\Sigma\beta_i\lambda_i$ as we would by considering the system as a single unit.

The addition of components to the system none of which have failed, i.e., data of the type $(t_i,0)$, do not necessarily cause the confidence to go rapidly to zero. (Of course, the confidence does depend upon t_i through θ.) It is clear from (5.5) that it is not the number of components but the number of failures which rapidly decrease the confidence.

In the special case when $\beta_{ijk} = \beta_i$ for all j,k we can make an intuitive interpretation of γ_i as the conditional probability of failure of the i^{th} assembly given that an assembly has failed. To see this, label the events "the i^{th} assembly fails" by F_i and "the i^{th} assembly malfunctions" by M_i. By definition

$$\beta_i = P[F_i|M_i], \qquad \alpha_i = P[M_i|\Sigma M_j]$$

and from the calculus of probabilities, since $F_i \subset M_i$

$$\alpha_i \beta_i = \frac{P(F_i)}{P[\Sigma M_j]}$$

and hence from (5.3.1) follows $\gamma_i = P[F_i | \Sigma F_i]$.

We now make two calculations to indicate the reliability values obtained by this method.

Example 1:

Let us suppose that $\gamma_i = \frac{1}{m}$ for $i=1,\ldots,m$. We recall that under certain conditions this would mean the event any one particular component had failed, knowing that exactly one component was in a failed state, was equally likely with the event any other component had failed.

From Table I we find $k = 8$ and compute from (9.4), $\Theta = \Sigma\tau_i/m = 25.35$ and hence for $\epsilon = .95$, using the Chi-square value for 16 degrees of freedom, we have $u = (28.87)/50.7 = .569$. Thus $e^{-u} = .566$ is a lower 95% confidence limit for the system reliability.

Example 2:

Let us suppose $\alpha_i = \frac{1}{m}$ for $i=1,\ldots,m$ and from Table I, we again use (9.4) to compute $\Theta = (\Sigma t_i)/(\Sigma \beta_i) = 16.27$. For $k = 8$, $\epsilon = .95$ we find $u = (28.87)/(32.54) = .887$ and $e^{-u} = .412$ is the lower 95% confidence limit for the reliability of the system.

6. BOUNDS ON θ

In this section we make the argument that what prior information one has about λ_i for $i=1,\ldots,m$ should be applied so as to determine bounds on θ rather than in the production of prior distributions of the component failure rates.

It is clear that if $\underline{\alpha} = (\alpha_1,\ldots,\alpha_m)$ is constrained and a lower bound $\theta_1 \leq \theta(\underline{\alpha})$ can be determined, then correspondingly from (5.7) $u_1 \geq u$, from which it follows that e^{-u_1} provides a lower confidence bound of level not less than ε. For example, the trivial inequality

$$\min_{i=1}^{m} \tau_i \leq \theta(\underline{\alpha})$$

will provide such a bound. However, unless the τ_i are nearly all equal, a state devoutly to be wished and planned for, it is not certain this bound would be a useful result. However, if testing were continued until $\tau_i = \tau_0$ for $i=1,\ldots,m$, we would then be in the favorable position that α_i need not be known.

If τ_j were the minimum of τ_i for $i=1,\ldots,m$, then $\tau_j = \theta(\underline{\alpha})$ implies $\gamma_j = 1$, $\gamma_i = 0$ for $i \neq j$ which in turn by (5.3.1) implies that $\alpha_j = 1$, $\alpha_i = 0$ for $i \neq j$ which requires by (5.0.1) that λ_j be infinitely large with respect to all other λ_i. This would seem to be an unlikely state of nature, one which might be reasonably excluded from consideration.

We now give some examples of information which in various degrees exclude the state mentioned above and are of a type which may provide a non-trivial bound.

Let us suppose it is known that

$$\sum_{i=1}^{m} \alpha_i \beta_i = p. \tag{6.1}$$

(This would mean in certain situations that the probability of a failure given a malfunction was known to be p.)

Thus the problem becomes that of minimizing the function $\theta(\underset{\sim}{\alpha})$ as defined in (5.4) where (t_i, β_i) are known positive numbers subject to the restriction (6.1) and

$$\sum_{1}^{m} \alpha_i = 1, \quad \alpha_i \geq 0 \quad \text{for} \quad i=1,\dots,m. \tag{6.2}$$

Call the set of $\underset{\sim}{\alpha}$ satisfying (6.1) and (6.2) the set $\mathcal{A}_p$.

Clearly with the denominator fixed in (5.4) we have a linear programming problem with two constraints, for which the theory is well known.

Of course the restriction (6.1) is a mathematical convenience. What we desire are bounds on $\min\limits_{p} \psi(p)$ for p taken over some subset of the range

$$\min \beta_i < p < \max \beta_i,$$

where

$$\psi(p) = \min\{\theta(\underset{\sim}{\alpha}) : \underset{\sim}{\alpha} \in \mathcal{A}_p\}. \tag{6.3}$$

This can be obtained from a graph of $\psi(p)$, which is here accomplished with a linear program using p as a parameter. A plot using the data of Table I is given in Figure 2 as on illustration.

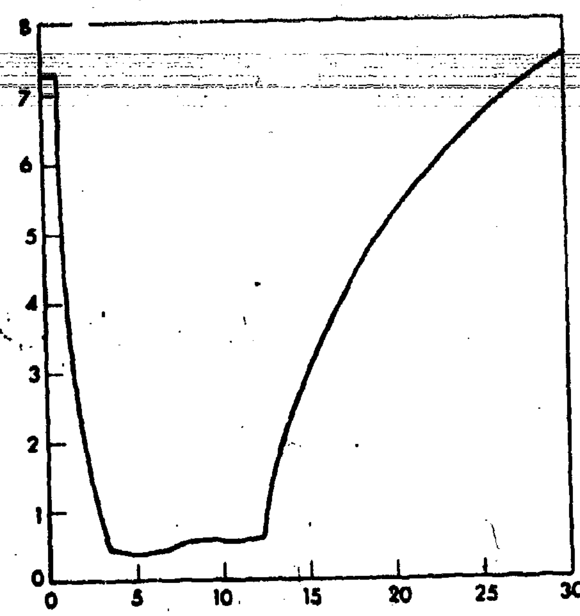


Figure 2

Graph of $\psi(p)$ for $1 \leq p \leq 30$.

Consider θ as a function of $\underset{\sim}{\gamma} = (\gamma_1,\ldots,\gamma_m)$ as defined in (5.4). It is linear (hence convex) and we wish to minimize it subject to some restrictions on its domain which are measures of the variability of the λ_i for $i=1,\ldots,m$. First we shall consider the region $\mathcal{B}_x$ defined for $0 \leq x \leq \sqrt{m-1}$ by

$$(6.4) \qquad 1 \leq m\Sigma\gamma_i^2 \leq x^2+1, \qquad \Sigma\gamma_i = 1, \qquad \gamma_i \geq 0 \quad \text{for} \quad i=1,\ldots,m.$$

The region $\mathcal{B}_x$ is convex and thus there exists a unique minimum for θ over the region. The method we shall use is Lagrange multipliers.

Let

$$\phi(\underset{\sim}{\gamma}) = \theta(\underset{\sim}{\gamma}) - \frac{m\lambda_1}{2}\Sigma\gamma_i^2 - \lambda_2\Sigma\gamma_i.$$

We wish to minimize ϕ subject to the conditions

$$(6.5) \qquad m\Sigma\gamma_i^2 = x^2+1, \qquad \Sigma\gamma_i = 1, \qquad \gamma_i \geq 0 \quad \text{for} \quad i=1,\ldots,m.$$

Thus

$$\frac{\partial\phi}{\partial\gamma_j} = \tau_j - m\lambda_1\gamma_j - \lambda_2 \qquad j=1,\ldots,m.$$

We now consider the three equations

$$(6.6) \qquad \Sigma\gamma_i\frac{\partial\phi}{\partial\gamma_i} = 0, \qquad \frac{1}{m}\Sigma\tau_i\frac{\partial\phi}{\partial\gamma_i} = 0, \qquad \frac{1}{m}\Sigma\frac{\partial\phi}{\partial\gamma_i} = 0$$

which upon simplication, and imposing the restrictions of (6.5), yield three equations which are to be solved for θ by eliminating λ_1 and λ_2

First eliminating λ_2 we obtain

$$(6.7) \qquad \lambda_1 = \frac{\sigma^2}{\theta-\bar{\tau}}, \qquad (\theta-\bar{\tau})^2 = x^2\sigma^2$$

where $\bar{\tau}$ and σ are the mean and standard deviation of τ_i's, respectively.

Thus for a given value of x we can obtain the minimum value of θ from (6.7) as the smaller root

$$(6.8) \qquad \theta = \bar{\tau} - x\sigma.$$

Also from $\frac{\partial\phi}{\partial\gamma_j} = 0$ we must have

$$(6.9) \qquad \gamma_j = \lambda_1^{-1}m(\tau_j-\lambda_2) \geq 0 \qquad \text{for} \qquad j=1,\ldots,m$$

in order to satisfy the restrictions. Since by (6.7) and (6.8) $\lambda_1 = -\sigma/x$ we see that a sufficient condition to satisfy (6.9) is

$$(6.10) \qquad \max(\tau_j) \leq \lambda_2 = \bar{\tau} + \sigma x^{-1}.$$

This is satisfied for all τ_j which are reasonably close together. It is clear from (6.8) that the minimum value of θ is a decreasing function of x.

<u>Remark</u>: If $\tau_1,\ldots,\tau_m$ satisfy (6.10) for a given x between $0 \leq x \leq \sqrt{m-1}$, then

$$\min\{\theta(\gamma): \gamma \in \mathcal{B}_x\} = \bar{\tau} - x\sigma$$

where $\bar{\tau},\sigma$ are defined in (6.7).

From this remark we see the restriction (6.4) yields a lower $100\varepsilon\%$ confidence bound for the reliability, namely

(6.10.1) $\exp\{-\lambda_\varepsilon^2(2k+2)/2(\bar{\tau}-x\sigma)\}$ for $0 \le x \le \sigma/(\max \tau_j-\bar{\tau})$.

Using the data in Table I we find $\bar{\tau} = 25.35$, $\sigma = 21.25$ and $\max \tau_j = 98.1$ and thus the range of x is $0 \le x \le .31$. A graph of (6.10.1) for this case with $k = 7$, $\varepsilon = .95$ is given in Figure 3.

Next we consider the problem of minimizing $\theta(\underset{\sim}{\gamma})$ subject to

(6.11) $\Sigma\alpha_i^2 = \kappa$, $\Sigma\alpha_i = 1$, $\alpha_i \ge 0$ for $i=1,\dots,m$.

Since we can write

$$\alpha_j = \gamma_j\beta_j^{-1}/(\Sigma\gamma_i\beta_i^{-1}) \quad \text{for} \quad j=1,\dots,m,$$

the first restriction can be put in the alternate form $\Sigma\gamma_j^2\beta_j^{-2} = \kappa(\Sigma\gamma_j\beta_j^{-1})^2$. Again we use Lagrange multipliers to take advantage of the symmetry of the problem. Write

$$\phi = \theta - \frac{m\lambda_1}{2}[\Sigma\gamma_i^2\beta_i^{-2}-\kappa(\Sigma\gamma_i\beta_i^{-1})^2] - \lambda_2(\Sigma\gamma_i-1)$$

$$\frac{\partial\phi}{\partial\gamma_j} = \tau_j - m\lambda_1[\gamma_j\beta_j^{-2}-\kappa(\Sigma\gamma_i\beta_i^{-1})\beta_j^{-1}] - \lambda_2.$$

We now consider the four equations

$$\frac{1}{m}\sum_1^m \beta_j\frac{\partial\phi}{\partial\gamma_j} = 0,\quad \frac{1}{m}\sum_1^m \beta_j^2\frac{\partial\phi}{\partial\gamma_j} = 0,\quad \sum_1^m \gamma_j\frac{\partial\phi}{\partial\gamma_j} = 0,\quad \sum_1^m \tau_j\beta_j^2\frac{\partial\phi}{\partial\gamma_j} = 0.$$

By setting $\delta = \Sigma\gamma_j\beta_j^{-1} = (\Sigma\alpha_i\beta_i)^{-1}$ for notational convenience and imposing the restraints as encountered we obtain four equations which contain the four variables $\lambda_1,\lambda_2,\theta,\delta$. Eliminating λ_1,λ_2 and δ gives a quadratic equation in θ, namely

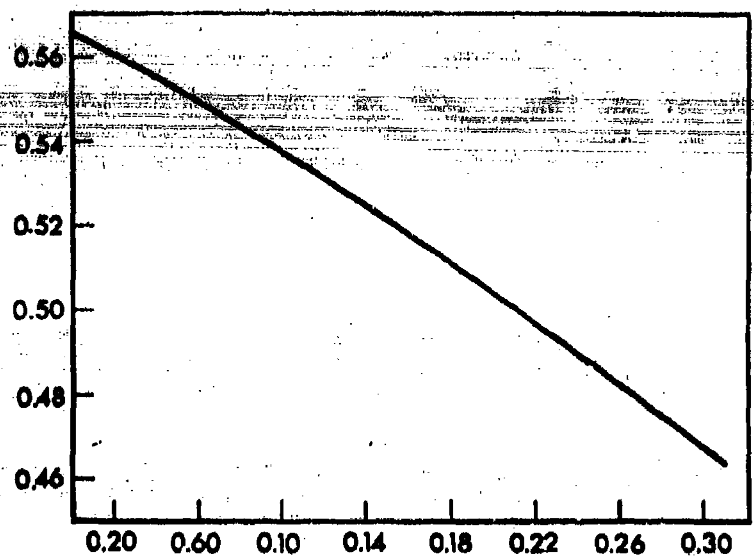


Figure 3

Values of $\exp\{-28.87/2(25.35-21.25x)\}$ for $0 \le x \le .31$: a lower 95% confidence bound on the reliability given $1/m \le \Sigma\gamma_i^2 \le \frac{x^2+1}{m}$.

$$\theta^2 A - 2\theta B + C = 0$$

where, recalling $t_i = \beta_i \tau_i$ for $i = 1, \ldots, m$,

$$A = (\bar{\beta})^2 - x\overline{\beta^2}, \qquad B = \bar{t}\bar{\beta} - x\overline{t\beta}, \qquad C = (\bar{t})^2 - x\overline{t^2}$$

and all coefficients depend upon the value of $x = 1 - (m\kappa)^{-1}$. Thus the minimum value is the smaller root, call it

(6.14) $$\theta_x = (B_x - S_x)/A_x$$

where $S^2 = B^2 - AC$ or equivalently $S_x^2 = xa^2 - x^2 b^2$

and

$$a^2 = \frac{1}{m}\Sigma(t_i\bar{\beta} - \beta_i\bar{t})^2, \qquad b^2 = \overline{t^2}\,\overline{\beta^2} - (\overline{t\beta})^2.$$

It is clear that if x is constrained by

(6.14.1) $$0 \le x \le \min\left\{\frac{(\bar{\beta})^2}{\overline{\beta^2}}, \frac{\bar{t}\bar{\beta}}{\overline{t\beta}}, \frac{(\bar{t})^2}{\overline{t^2}}, \frac{a^2}{2b^2}, 1 - \frac{1}{m}\right\},$$

then the values of θ_x are meaningful.

We now argue that there is an interval of positive values of x for which θ_x is decreasing as a function of x. Note $\theta_x' \le 0$ iff

(6.15) $$\zeta(x) = -S_x' A_x - \overline{\beta^2} S_x \le \overline{t\beta}\,(\bar{\beta})^2 - \bar{t}\bar{\beta}\overline{\beta^2}$$

To see that ζ is an increasing function, note that $\zeta'(x) = -A_x S_x''$. Now $S'(x) = (a^2-2xb^2)/2S(x)$ so that by (6.14.1), $S'(x) > 0$ and likewise we check $S''(x) < 0$ and hence $\zeta'(x) \geq 0$. Notice that $\zeta(0) = -\infty$ so there is an interval of values in x, for which (6.15) is true, and θ_x is decreasing. This region can be determined from (6.15) in each case of interest.

However, we must also satisfy the condition $\gamma_j \geq 0$ for $j=1,\ldots,m$. From $\frac{\partial\phi}{\partial\gamma_j} = 0$ follows, by using (6.13)

$$\text{(6.16)} \qquad \gamma_j \geq 0 \quad \text{iff} \quad 1 \geq \frac{\beta_j\theta - t_j}{\kappa m\lambda_1\delta} = \frac{x(t_j-\beta_j(\theta))}{\bar{t}-\bar{\beta}\theta}.$$

But by the argument above $\theta_x < \theta_0 = \bar{t}/\bar{\beta}$ for $x > 0$. Hence the denominator of the right-hand side of (6.16) is positive. Thus $\gamma_j \geq 0$ for $j=1,\ldots,m$ iff

$$\text{(6.17)} \qquad \bar{t} - \theta_x\bar{\beta} \geq x \max_{j=1}^{m}\{t_j - \theta_x\beta_j\}.$$

To be presuaded that there is indeed a neighborhood of zero in which (6.17) is true we introduce the power series expansion of θ_x:

$$\theta_x = \frac{\bar{t}}{\bar{\beta}} - \sqrt{x}\,\frac{a}{(\bar{\beta})^2} + O(x).$$

Thus (6.17) is equivalent with

$$\frac{a}{\bar{\beta}} - \sqrt{x}\,\frac{\bar{t}\overline{\beta^2} - \bar{\beta}\,\overline{t\beta}}{(\bar{\beta})^2} + O(x) \geq \sqrt{x}\max_j[t_j - \theta_x\beta_j]$$

which is clearly true for x sufficiently small. Thus we can make the

Remark: If $(t_i, \hat{\beta}_i)$ $i=1,\ldots,m$ are given and x satisfies (6.14.1), then θ_x defined by (6.14) satisfies

$$\theta_x = \min\{\theta(\underset{\sim}{\gamma}): \underset{\sim}{\gamma} \ni \frac{1}{m} \le \Sigma\alpha_i^2 \le \frac{1}{m(1-x)}\}$$

whenever (6.17) is true.

Using the data from Table I we find that the upper bound for (6.14.1) is $a^2/2b^2 = .26056$ while the largest value of x satisfying (6.17) is .145. A graph of $\exp\{-\chi_\varepsilon^2(2k+2)/2\theta_x\}$ for $k = 7$, $\varepsilon = .95$ with θ_x as defined in (6.14) is given in Figure 4.

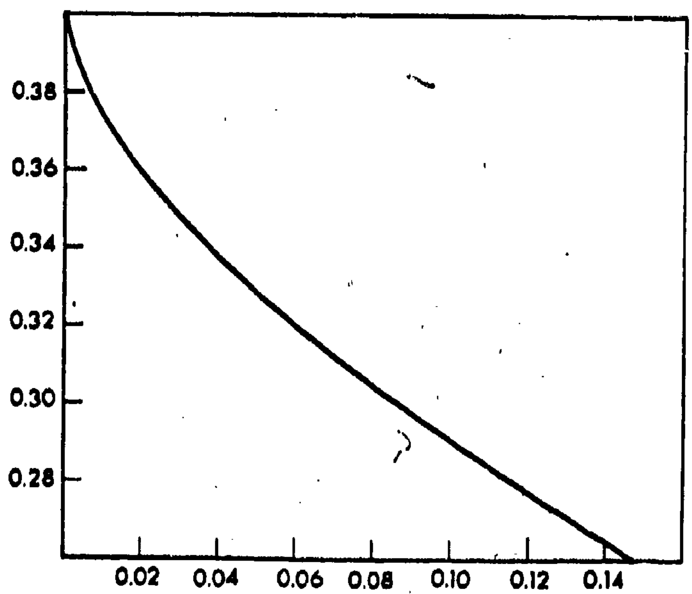


Figure 4

Graph of $\exp\{-\frac{28.87}{2\theta_x}\}$ for $0 < x < .15$: a lower 95% confidence bound on the reliability given $\frac{1}{m} \le \Sigma\alpha_1^2 \le \frac{1}{m(1-x)}$.

CONCLUDING REMARKS. By considering a series system with a given number of components, each component having been tested for the same time and each experiencing no failures, one intuitively feels since the entire system could have been operated, conceptually at least, without failure for that same period of time that the confidence one has in the system's reliability should be exactly the same as the confidence in each component's reliability. Moreover, this confidence should be the same irrespective of the number of components in the system.

By thinking how the confidence should behave for such a series system, as components are added with different test times and different numbers of failures, we see that the Bayesian approach with independent prior distributions of the failure rate fails to fulfill our expectation as to this incremental behavior. At the same time the second model chosen, with failure rates as virtual constants, does seem to behave in conformity with our intuition and moreover it has the added appeal of computational simplicity.

Lastly, for the practical case chosen namely early data for the Saturn 1-C which at this juncture we know is a highly reliable system, the second model gave reasonable interval estimates of the reliability while the first did not.

ACKNOWLEDGMENTS. The author would like to express his appreciation to the reliability group at the Launch Systems Branch of The Boeing Company for providing the data for this discussion and assistance in its organization. In particular, I want to thank Francis Bari who first called this question to my attention by mentioning some deficiencies of the extant methods for determining confidence bounds.

TABLE I

Summary of test data for Saturn I-C

t_i = test time in mission lengths, n_i = number of failures observed

ω_i = q_i = component multiplicity, $\tau_i = t_i/\beta_i$

t_i	n_i	ω_i	τ_i
318.5	0	16	19.9
138.8	0	16	8.7
69.4	0	16	4.3
159.2	0	4	39.8
187.9	0	8	23.5
144.9	0	4	36.2
69.7	0	4	17.4
148.7	0	4	37.2
146.8	0	4	36.7
15.1	2	2	7.5
7.5	0	12	.63
120.7	0	4	30.2
113.1	0	3	37.7
98.1	0	1	98.1
92.8	0	1	92.8
9.0	0	1	9.0
97.9	0	13	7.6
87.3	0	4	21.8
26.4	0	1	26.4
83.7	0	4	20.9
14.1	0	1	14.1
41.5	0	1	41.5
7.5	0	2	3.8
15.1	0	3	5.0
11.3	0	2	5.7
29.3	0	5	5.9
82.9	0	1	82.9
7.5	1	1	7.5
20.8	0	2	10.4
52.8	0	1	52.8
51.8	1	1	51.8
65.3	0	1	65.3
66.2	0	1	66.2
65.2	0	2	32.6
48.7	0	1	48.7
33.9	0	1	33.9
30.2	0	1	30.2
45.7	0	1	45.7
36.5	0	1	36.5
50.6	0	2	28.3
45.2	0	2	22.6
22.6	0	2	11.3
37.7	0	2	18.8
49.9	0	1	49.9
34.3	0	1	34.3
37.7	0	1	37.7
11.3	0	1	11.3
15.1	0	2	7.5
226.1	0	30	7.5
30.3	0	10	3.0
32.0	0	2	16
179.0	0	8	22.4
32.5	2	1	32.5
75.4	0	10	7.5
191.9	0	8	24
34.6	0	2	17.3
17.7	1	1	17.7
73.6	0	4	18.4
88.6	0	4	22.2
8.0	1	1	8.0
1.6	0	3	.5
2.4	0	6	.4
18.9	0	1	18.9
11.1	0	1	11.1
14.9	0	1	14.9
13	0	1	13
7.3	0	1	7.3

BIBLIOGRAPHY

1. Allen, D. C., Carlson, C. H., Hubach, C. E. Procedure for Reliability Assessment with Confidence. Federal Electric Corporation Document No. AS-A-86-67, June 1967.

2. Dalton, R. E. An Evaluation of Methods for Construction of Confidence Limits for System Reliability. TRW Systems, Florida Operations, Contract AF 04(694)-806, June 1966.

3. Donnelly, Patrick P. Is Reliability Necessary? Ninth Annual West Coast Reliability Symposium, Western Periodicals Company, North Hollywood, California, pp. 61-65, February 1968.

4. Esary, J.D., Proschan, Frank, and Marshall, A. W. Some Reliability Applications of the Hazard Transform. BSRL Mathematical Note No. 605, May 1969; to appear in SIAM J. Appl. Math.

5. Feller, William. An Introduction to Probability Theory and Its Applications, Volume II, John Wiley and Sons, 1966.

6. Lindley, D. V. Probability and Statistics, Volume II, Inference Cambridge University Press, 1965.

7. Madansky, Albert. Approximate Confidence Limits for the Reliability of Series and Parallel Systems, Technometrics 4, 495-503, 1965.

8. Mann, Nancy R. Confidence Bounds for System Reliability from Subsystem Data: A Survey of the State of the Art, presented at the Fifteenth Conference on the Design of Experiments in Army Research, Development and Testing. (To be published.)

9. Myhre, J. M., Saunders, S. C. On Confidence Limits for the Reliability of Systems, Ann. Math. Statist. 39, 1463-1472, 1968.

10. Myhre, J. M., and Saunders, S. C. Comparison of Two Methods of Obtaining Approximate Confidence Intervals for System Reliability, Technometrics 10, 37-50, 1968.

11. Saunders, S. C. On the Determination of a Safe Life for Distributions Classified by Failure Rate, Technometrics 10, 361-377, 1968.

12. Springer, M. D. and Thompson, W. E. Bayesian Confidence Limits for Reliability of Redundant Systems When Tests are Terminated at First Failure, Technometrics 10, 29-36, 1968.

13. Zimmer, W. J., Breipohl, A. M., and Prairie, R. R. A Consideration of the Bayesian Approach to Reliability Evaluation, IEEE Transactions on Reliability, R-14, No. 2.

A PROBABILITY MODEL FOR THE ASSESSMENT OF HUMAN INCAPACITATION FROM PENETRATING MISSILE WOUNDS

William P. Johnson
William J. Bruchey, Jr.
Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

ABSTRACT. A mathematical model is proposed for the evaluation of the altered performance of one of the most complex systems known to man, himself. Probabilities are associated with a random fragment penetrating varying distances within the human body, striking a critical anatomical component and inflicting damage to the extent that the wound recipient would be unable to perform his assigned duties.

These probabilities are combined to determine the conditional probability of all events occurring, simultaneously, to a tactical soldier under battlefield conditions.

The remainder of this article has been reproduced photographically from the authors' manuscript.

Preceding page blank

I. INTRODUCTION

To evaluate the effectiveness of antipersonnel weapons it has become necessary to employ a quantitative casualty criterion. A suitable criterion for many purposes is the probability that the weapon will wound, fatally or severely, its intended target.

An old criterion of wounding power is the 58 foot pound rule. In its crudest form it states that missiles with less than 58 foot pounds of kinetic energy do not kill, that those with more than 58 foot pounds of kinetic energy do kill. This criterion was never intended to be more than a rough rule of thumb. Burns and Zuckerman[1*] made a more refined analysis in 1941 of the quantitative requirements for wounding, while Guerney[2] suggested that MV^3 was a more suitable criterion than the kinetic energy standard for wounding human targets.

Wound ballistics work carried out in this country during and since World War II has included studies of the relationship between the volumes of cavities formed in tissues and tissue simulants by missiles and the physical parameters describing the projectile on impact with the target. Such work has not yet provided numerical relations between the physical parameters of fragments and the probability that the wounded man will suffer any specified degree of impairment of his ability to function. For the evaluation of the antipersonnel weapons, there is needed a knowledge of the numerical probability that a man, struck by a projectile of specified characteristics, will thereafter be unable to perform the functions of his tactical role. For evaluation or design, the size of a wound is not directly employable as a useful criterion. So far, cavity studies have unfortunately led to little or no information suitable for the evaluation or design of fragmenting weapons.

The Burns and Zuckerman studies led to numerical probabilities of the type needed for evaluation. A later numerical study of a similar type was reported by McMillen and Gregg of the Princeton Department of Biology, in Missile Casualties Report No. 12, 6 Nov 1945, National Research Council, Division of Medical Sciences. McMillen and Gregg were concerned with wounds that they considered to be either fatal or severe.

---

*Superscript numerals refer to references.

These could be caused, they assumed, by projectiles reaching certain vulnerable regions inside the body after traversing protective layers of skin, soft tissue, and bone provided that the projectiles reach the vulnerable regions with velocities in excess of 750 m/sec. The thicknesses of the anatomical structures that had first to be traversed were determined from anatomical charts depicting cross sections of the body at approximately one inch intervals from head to feet. The velocities necessary to penetrate the various thicknesses of skin, tissue and bone were determined through in vivo experimentation. The selection of vulnerable regions was somewhat arbitrary, since no experiments were conducted to determine which regions were in fact of primary importance. The final results, which were therefore semi-theoretical in nature, were in the form of probabilities that random hits with steel balls on a human target would cause fatal or serious wounds. These probabilities were plotted against the striking energies of the balls, and were on the whole consistent with the earlier conclusions of Burns and Zuckerman.

A generalization of quite a different type was published by T. E. Sterne[3] in May 1951. In that study the experimental data employed by McMillen and Gregg were re-examined and re-interpreted. In addition, the calculations of McMillen and Gregg were repeated for randomly shaped fragments similar to bomb fragments instead of spheres. The temporary cavities caused by the fragments penetrating the target were taken into account by requiring that a fragment, in order to cause important injury to a vital region, must not only penetrate the intervening skin, tissue, and bone, but must also reach the vulnerable regions with sufficient remaining energy to produce a temporary cavity of 2 cubic centimeters.

In October of 1956, "A New Casualty Criterion for Wounding by Fragments" was published by Allen and Sperrazza.[4] This study revised the combination of mass and velocity combinations which appeared to be related to incapacitation. Instead of employing the MV/A parameter used by Sterne, MV^{β}, where $1 < \beta < 2$, was introduced for the first time. This criterion has been used extensively to determine the relationship between trauma and fragment characteristics. However, close examination of the techniques used within the current model reveals that the resultant quantitative data is based upon numerous subjective assumptions which vary from one evaluation to the next and that the resultant data are not probabilistic numbers in the mathematical sense.

With the advantage of the knowledge available in the area from earlier investigations, it is the purpose of the current study to establish a model that will supersede the currently used fragment criteria. The specific objectives of the proposed study are:

(1) To establish a truly probabilistic model for the assessment of human incapacitation from penetrating missile wounds.

(2) To eliminate the necessity for tissue retardation firings and in vivo experimentation as a prerequisite for the evaluation of the wounding potential of future fragments.

(3) To eliminate the need for trajectory tracings through anatomical cross sections of the human body in order to establish quantitative values for the probability of incapacitation.

(4) To establish a common basis for the comparison of all fragmenting munitions.

The approach and procedure for establishing the proposed model while accomplishing the above objectives, are presented in the following sections of this paper.

II. APPROACHES AND ASSUMPTIONS

For the purposes of establishing a mathematical model to predict the probability of incapacitation to a human upon impact from a random projectile the following assumptions are made:

(1) The fragmenting pattern of grenades and other exploding munitions have been studied in sufficient detail to provide the probability of a specific fragment striking a human target, P(H).

(2) Each fragment impacting the human target has a distinct probability of penetrating specific distances within the wound recipient's body, P(P).

(3) Each fragment striking the casualty has a distinct probability of encountering a critical organ along its path, P(E). This probability is conditioned by the penetrating ability of the fragment and the location of the critical components of the human anatomy.

(4) Each anatomical component has a distinct probability of being damaged, P(D), to the extent of preventing specific biomechanical motions required by a tactical soldier for the full performance of his mission.

(5) Each specific biomechanical motion has a distinct probability, P(M), of being required during the performance of the soldiers total mission.

(6) The probability of incapacitation from the N'th, P(N), component may be described by the expression:

$$P(N) = P(H) \cdot P(P/H) \cdot P(E/HP) \cdot P(D/HPE) \cdot P(M/HPED)$$

(7) Each of the component probabilities are independent and can be combined mathematically to provide the conditional probability of incapacitation, P(I/HPEDM), to a tactical soldier from a random projectile by the following:

$$P(I/HPEDM) = \frac{\sum_{1}^{N} P(N)}{N}$$

For the purpose of brevity, henceforth the conditional probability of incapacitation of a tactical soldier, P(I/HPEDM), will be shortened to P(I/H). It is to be understood that the second expression includes all of the conditions provided within the first expression.

(8) No synergistic effects occur from multiple wounds. Thus, the probability of incapacitation to a tactical soldier from two or more wounds can be determined by mathematically combining the independent conditional probabilities associated with each wound using the following expression:

$$P(I/H) = 1-(1-P(I/H_1))(1-P(I/H_2))(1-P(I/H_3)).....(1-P(I/H_N))$$

As mentioned earlier, the probability of a fragment striking a stationary human target is assumed to have been established by Exterior Ballistics experts. Therefore, the remainder of this paper will concentrate on the procedures to be used to establish the probabilities required for the other portions of the model.

III. PROBABILITY, P(P), THAT PROJECTILE WILL PENETRATE A SPECIFIC DISTANCE, (D), WITHIN THE HUMAN BODY

As assumed by McMillan and Gregg, it seems reasonable to suppose that the probability that a random hit will cause fatal or severe wounding depends upon the fraction of the body's superficial area through which the fragment can wound a vital organ. The identification of the vital regions, and the conditions of striking them necessary to cause fatal or severe wounding may not have been correctly chosen by McMillen and Gregg, nevertheless, it still seems reasonable to suppose that the probability P(I/H) will be a function of the penetrating ability of a fragment. If fragments possess such great penetrating power that they traverse a body completely wherever they hit, then the probability P(I/H) that a random hit will incapacitate is the ratio of a rather large vulnerable area to the total presented area of the body. On the other hand, if the penetrating power of a fragment is so low that it can never reach the critical components of the body, then the probability P(I/H) must approach zero.[5]

It is proposed that an indication of the penetrating potential of fragments be obtained thru the use of facilities currently or soon to be available to wound ballisticians. These include (1) the Ballistic Research Laboratories Computer Man[6] and (2) striking versus residual velocity comparisons for each of the anatomical components of the human body.

A brief diversion is required at this point to acquaint the reader with the BRL Computer Man, in order that he may fully appreciate its potential for aiding in the solution of the current problem.

In brief, the BRL Computer Man is a computer program currently used within the wound ballistics program to determine the extent of incapacitation experienced by wounded tactical soldiers. It consists of coded versions of human anatomical cross sections extracted from Eycleshymer and Schoemaker, "A Cross Section Anatomy." The combination of the cross sections represents a human male in a specified tactical position. Every major anatomical component illustrated in the original cross section anatomy has been coded within the computer version in approximately the same proportion as found in the published version.

The associated instructions accompanying the coded cross sections within the computer model permit simulated fragment paths to be traced through the individual cross sections at various impact angles. Retardation data is provided for each of the anatomical structures and is used to determine the velocity loss within each coded component as a function of the distance traversed by the fragment through the individual components. Thus the penetrating ability of each fragment becomes a function of (1) striking conditions at impact upon the cross section and (2) the retarding ability of the anatomical components encountered along its path.

Retardation data for several anatomical components has been used within the Wound Ballistics program for several years. However, the data in existence is of inadequate quality for the ultimate solution to our current problem. Fortunately, a program is currently being conducted by Sturdivan and Thompson[6] which should provide the necessary retardation data input to the BRL Computer Man.

Upon completion of the retardation studies, the data will be fed into the Computer Man model. Maximum distances traveled within the anatomical model will be determined as a function of several physical parameters of the impinging projectile (mass, velocity, presented area). Success or failure for each trajectory will be determined by its ability to penetrate at least each of several pre-specified distances within the human anatomy. The probability, P(G), of a particular fragment-velocity combination penetrating at least each of the pre-specified distances within an anatomical subdivision will be the ratio of the total number of successes at that distance to the total number of initiated trajectories not perforating the subdivision. It should be noted that the denominator becomes the number of non-perforating trajectories rather than the total number of trajectories originated within a subdivision. This becomes necessary because of the varying distances within the human anatomy at which perforations will be achieved by most projectiles because of the variations in tissue structures encountered along each path. In addition, the geometry of the human anatomy is such that the distance available for penetration by any projectile becomes smaller as one moves farther from the center of the body.

If we let P' represent the probability of a projectile perforating the human anatomy, then (1-P') represents the probability that the projectile will stop within the human body. However, it is desired that probability figures be determined as a function of several distances within the non-perforating group. The probability, P(P), associated with each non-perforating projectile penetrating at least each of the prespecified distances will be given by the product of the probability of retention of the fragment within the human anatomy, (1-P'), and the probability of the fragment reaching the required distance, P(G).

Symbolically, this becomes: $P(P) = (1-P') \cdot P(G)$

There appears to be two alternatives for considering the perforating projectiles:

(1) It can be assumed that the penetration pattern of the non-perforating projectiles are representative of that which would have been displayed by the perforating projectiles, if the human anatomy was such that it allowed all projectiles to penetrate as deeply as possible. Based on this assumption, the penetration pattern as a function of distance would be represented by P(P) and the perforating projectile paths would only be used to condition the probability of the body retaining an impinging projectile.

(2) It can be assumed that the perforating fragment paths represent the upper limits of penetration and that because they do perforate, the fragments would be able to reach any desired depth within the human anatomy. Based on this assumption, the penetration pattern as a function of distance would be represented by the sum of P(P) and P'.

As of this writing it has not been decided which of the two alternatives will be used within the proposed model. However, since we are aware of the available options, it is felt that subsequent discussions with knowledgeable personnel will provide insight as to the proper approach.

Probability data of the above type will be obtained for several missile-velocity combinations. It is proposed that regression equations be used to relate the probability of penetration of specific missiles to a physical parameter of the projectile. Thus, generalized equations can be developed which will provide the user with the probability of a particular projectile penetrating a specified distance within a human as a function of its impact point on the anatomy and other relevant physical characteristics of the impinging projectile.

IV. PROBABILITY OF PROJECTILE ENCOUNTERING THE N'TH ANATOMICAL COMPONENT DURING ITS PATH THROUGH THE HUMAN BODY, P(E)

If it is assumed that a projectile will impact the human target randomly, then it must be assumed that the path of the projectile through the body will be random. Although the proposed model is designed to yield primarily, the probability of incapacitation from a single penetrating projectile, it should be realized that an almost infinite combination of organs or tissues can be encountered along the path of a single wound. For this reason one must deal with the probability of encountering each independent anatomical component within a given wound tract, rather than combinations of traumatized organs. The shielding effects afforded some organs by their surroundings, combined with the orientation of the organs to specific impact angles makes the ratio of their size to the total body area an inaccurate measure of the probability of encountering the organ.

In addition, because the geometry of most anatomical components vary as a function of their depth within the anatomy, probabilities of encountering each structure must be developed as a function of the obliquity angle of the impinging projectile and the target depth within the human body. However, for the purposes of this model, final probabilities will be documented as a function of penetration distance only. Several obliquity angles will be used to determine the probability of encountering each structure as a function of its depth within the human anatomy . The final probability figures will reflect the weighted averages obtained from each of the independent analyses. Not only will this approach reduce the amount of bookkeeping involved within the project, it should also yield probabilities of encountering specific organs more representative of those expected from a truly random penetration.

The BRL Computer Man model will serve as the basis for the data obtained within this phase of the model. Imaginary trajectories will be traced through the coded cross sections and allowed to penetrate specific distances within the human anatomy. For each organ or tissue under consideration, a strike upon the structure within a specified distance will be recorded as a success. The probability of encountering each structure as a function of the specified distance will be the ratio of the total number of recorded successes to the total number of trajectories traced through the anatomy. The above will be documented at unit intervals for each anatomical component. The final probabilistic numbers will reflect the chances of having encountered each component with penetrations up to the indicated depth.

Subsequent additions to the proposed model will provide insight on the effects of missile "bite" on the probability of encountering components within the human anatomy. This physical phenomena considers the fact that a projectile need not actually strike a component in order to damage it. Investigations in these areas are currently underway within the Wound Ballistics program and once these data are available in quantitative form, the proposed model can be modified to include this phenomena.

Subsequent acquisition of modern anatomical cross sections reflecting the internal structure sizes of soldiers from specific military anthropometric percentiles should allow the probabilistic data generated during this subportion of the model to be generalized as a function of a physical measure of the human anatomy. No data exists of this type at the present time, therefore it is impossible to make such generalizations or even speculate as to how the probabilities of organ encounter are expected to vary between population percentiles.

V. PROBABILITY, P(D), THAT DAMAGE TO AN ANATOMICAL COMPONENT WILL PREVENT THE PERFORMANCE OF SPECIFIC BIOMECHANICAL MOTIONS

Before a missile wound can be evaluated in terms of the biomechanics required to perform a given task or role, it is necessary to describe the wound in terms which can be easily related to the physiological function of the body or subsystems within the body. That is, criteria must be developed which relates the probability of producing a given damage level to performance. This must be carried out in three steps. They are:

(1) Determine criteria to be used for damage description of the anatomical components.

(2) Determine the probability that missiles from a given munition will produce a specific damage level to a particular anatomical component.

(3) Determine the probability that this damage level will result in some biomechanical decrement.

At present work is being conducted under area (1). This task is oriented towards wound description in terms which can be related to biomechanical decrement. The criteria used to describe the wound will depend on the particular tissue or tissue type encountered. For instance, injury to muscles may be described in terms of hole size, percent muscle severed, etc; injury to blood vessels may be described in terms of the rate of blood flow from the injured location. Once these criteria have been determined, step (2) can be undertaken.

Presently, there exists a data bank of wound descriptions through animal tissue in terms of various parameters for several penetrating missiles. If, for instance, the criteria chosen is hole size, the wound data will be used to determine the distribution of hole size as a function of the missile parameters (mass, presented area, velocity). The information combined with the distribution of missile parameters for a given munition will provide the distribution of hole size for that munition. Using the probability distribution for hole size for a given munition an expected hole size will be computed, i.e. the expected damage level, E(D). Likewise, if the criteria for damage was blood loss, a similar E(D) would be computed.

This value of E(D) would then be presented to a medical assessor who would subjectively determine the probability that the indicated damage level would prevent an individual from accomplishing a specific biomechanical motion.

It is recognized that this subjective input into the model prevents the final results from realizing complete objectivity. However, in order to quantify the effects of damage levels to various anatomical components, years and years of experimental laboratory work would be required. As of this writing little experimental work has begun. In view of the fact that the decision made by the medical assessors will be a function of the expected damage to an anatomical component, totally independent of the physical parameters of the impinging projectile and made only once within the lifetime of the model, (except perhaps to improve a previous decision), it is felt that the subjectivity will be minimized as much as possible.

VI. PROBABILITY, P(M), THAT A SPECIFIC BIOMECHANICAL MOTION WILL BE REQUIRED DURING THE PERFORMANCE OF A SOLDIER'S MISSION

Due to the wide variety of specialties among servicemen, it is almost impossible to generalize upon the duties of active combatants. However, basic to the performance of duties associated with every specialty is the ability of the individual to perform controlled movements or biomechanical motions. Regardless of the traumatized anatomical components or region, a soldiers ability to function under battlefield conditions is directly related to his ability to carry out controlled movements. If an individual's wound site is such that it does not hamper or impair his ability to make the necessary biomechanical motions, then for all practical purposes, the individual cannot be regarded as incapacitated. These controlled movements of the body may be resolved into the functioning of the following subsections:

I. Shoulder Girdle

II. Shoulder Joint

III. Elbow and Radio-Ulnar Joint

IV. Wrist and Hand

V. Pelvic Girdle and Hip Joint

VI. Knee Joint

VII. Ankle and Foot

VIII. Spinal Column

IX. Thorax

I. Movement of the Shoulder Girdle:

1. Adduction - movement of the scapula medially toward the spinal column.
2. Abduction - sliding of the scapula laterally and forward along the surface of the ribs.
3. Elevation/depression - the upward and downward motions of the whole scapula without any rotation.
4. Upward rotation - involves an upward turning of the glenoid cavity and the lateral angle in relation to the superior angle and medial border, which turn downward.
5. Downward rotation - reverse of upward rotation.
6. Foreward tilt - occurs when the inferior angle moves backward away from the rib cage.
7. Backward tilt - the inferior angle and the costal surface return to the surface of the rib cage.

II. Movement of the Shoulder Joint:

1. Flexion - foreward elevation of the arm.
2. Extension - return movement.
3. Abduction - sideward elevation of the arm.
4. Adduction - return movement.
5. Inward rotation - turning the humerous around its long axis so that its anterior aspect moves medially.

6. Outward rotation - the opposite with the anterior aspect moving laterally.

III. Movement of Elbow and Radio-Ulnar Joints:

A. Elbow Joint:

1. Flexion.
2. Extension.

B. Radio-Ulnar:

1. Pronation.
2. Supination.

IV. Movements of the Wrist and Hand:

A. Wrist Joint:

1. Abduction.
2. Adduction.
3. Circumduction.
4. Flexion.
5. Extension.

B. Hand:

1. Prehensile movement.
   a. power grip - an object is clamped by the partly flexed fingers and palm with counter pressure applied to the thumb lying more or less in the plane of the palm.
   b. precision grip - object is pinched between the fingers and the opposing thumb.
2. Nonprehensile movement - objects are manipulated by pushing or lifting.

V. Movement of the Pelvis Girdle and Hip-Joint:

A. Pelvis:

1. Forward rotation - increased inclination resulting from lumbo-sacral hyperextention.
2. Backward rotation - opposite movement.
3. Lateral tilt - the lowering or raising of one iliac crest.
4. Rotation - turning about a vertical axis either to the right or left.

B. Hip-Joint:

1. Flexion - forward movement of the femur.
2. Extension - reverse movement.
3. Abduction - movement of one limb away from the other toward the side.
4. Adduction - movement of one limb from the side towards the other.
5. Circumduction - movement of the limb in a circular manner. Combine movements 1 - 4.
6. Rotation - may be outward or inward depending on which way the toes are turned.

VI. Movements of the Knee Joint:

1. Flexion.
2. Extension.
3. Inward rotation.
4. Outward rotation.

VII. Movements of the Ankle and Foot:

A. Ankle and Foot:

1. Dorsiflexion - consists of raising the foot toward the anterior surface of the leg.
2. Plantar flexion - lowering the foot so as to bring its long axis in line with that of the leg.
3. Eversion - the sole is turned laterally or outward.
4. Inversion - sole is turned medially.

B. Toes:

1. Flexion.
2. Extension.

VIII. Movements of the Spinal Column:

A. Cervical Spine:

1. Flexion.
2. Extension.
3. Lateral flexion.
4. Rotation.

B. Thoracic and Lumbar Spines:

1. Flexion.
2. Extension.
3. Lateral flexion.
4. Rotation.

IX. Movements of the Thorax:

1. Elevation.
2. Depression.

As can be seen from the above the possibility of a wide variety of body movements exists for every individual. However, we are concerned only with the essential or necessary movements involved in the tasks of combatants.

It is assumed that every task, regardless of its complexity, can be resolved into a series of controlled movements of the type presented above. It is further assumed that a probabilistic figure which reflects the chances of an individual being required to perform a specific biomechanical motion during the course of his duties can be determined. If an individual is required to assume multiple duties, then the probability of his performing a specific biomechanical motion will be conditioned by the probability of his performing the duty which requires the motion. Thus, the evaluation of a soldier's incapacitation can be related to several duties required within an overall mission.

In order to obtain the desired probabilities, one must be knowledgeable of the specific duties required of today's soldier. In addition, these duties must be analyzed from the kinesiologist point of view in order to reduce them to the series of independent biomechanical motions required for probabilistic analysis.

Once the duties have been reduced to a series of controlled movements, two alternatives for determining the probability, P(M), that a specific biomechanical motion will be required during the performance of a soldier's duties or mission appear to exist. The first assumes that each motion is independent of time and that every motion involved within a specific task requires the same percentage of the soldier's time. This does not eliminate the possibility of repetitions of the same movement within a given task. Repetitions are accounted for by including each repetition as a separate component of the total sum of movements involved. The probability, P(M), of a particular motion being required within the task will be the ratio of the total repetitions of the movement within the task to the total number of independent movements involved.

The second approach considers the total time required by the combatant to perform his duties as the unit and the probability, P(M), is determined by the functional part of the total allocated to each independent motion.

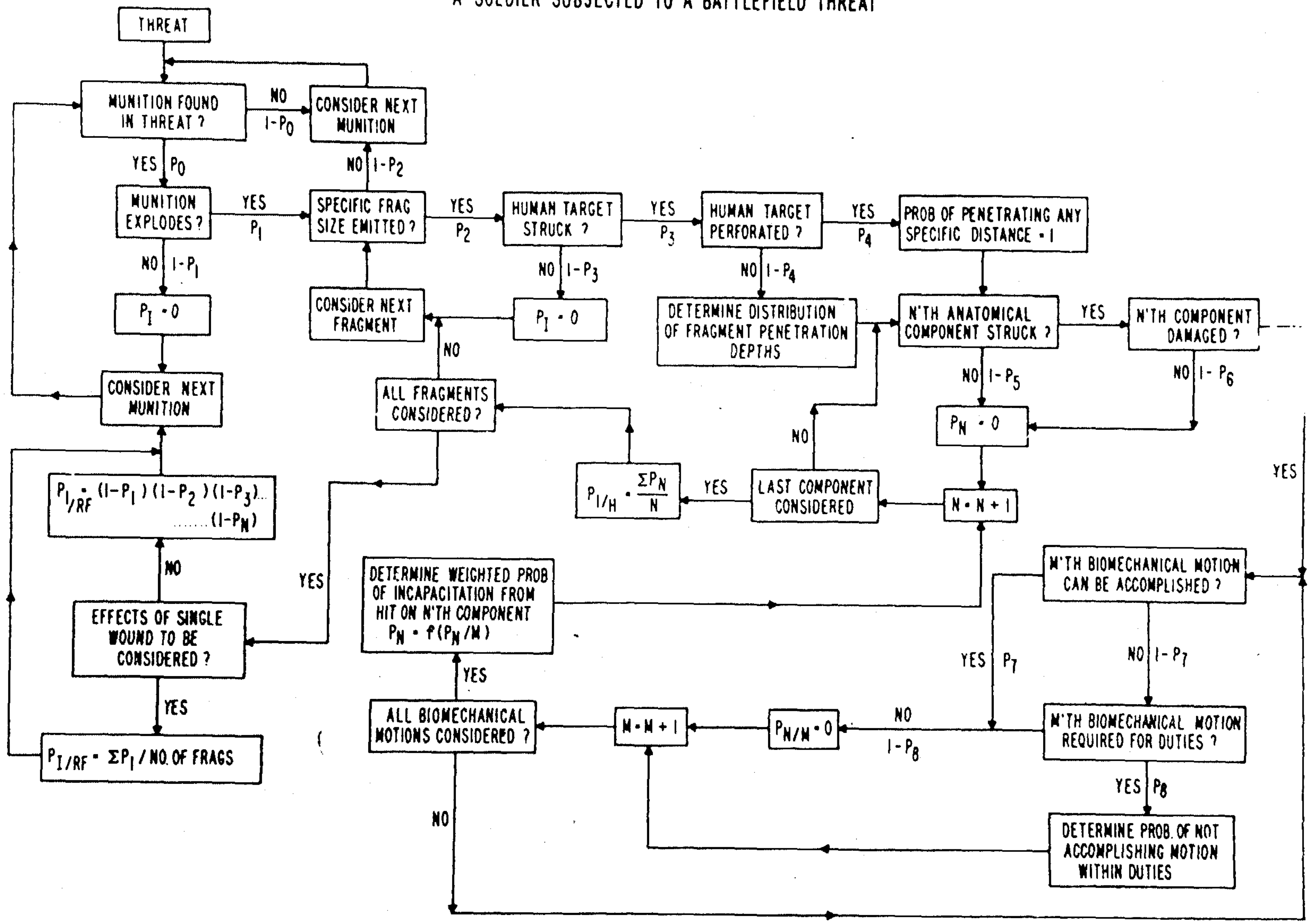
Each of the two approaches has advantages and disadvantages obvious to the authors at this point. However, suffice it to say that the merits of either approach will be discussed in detail with the experts on the subject prior to deciding which approach is most feasible. It appears that either of the alternatives will yield probabilistic data of the type required for input into the model. Therefore, no serious problems appear to exist.

VII. SUMMARY

A mathematical model for assessing the probability of human incapacitation from a penetrating missile wound has been proposed. The model assumes that several probabilistic events must occur, sequentially, in order to cause incapacitation to a tactical soldier. Among these events are (1) the soldier must be hit, (2) the impacting projectile must penetrate deep enough into the human anatomy to strike a critical anatomical component, (3) the anatomical component must be damaged to the extent that its physiological function is impaired, and (4) the dysfunction of the component must be directly related to the soldier's ability to perform specific biomechanical motions required for the performance of the soldier's task.

An approach has been given for obtaining the necessary experimental data needed to relate each of the above events in a probabilistic manner. In addition, the assumptions and anatomical equations used within the model have been detailed.

LOGIC FOR DETERMINING THE PROBABILITY OF INCAPACITATING
A SOLDIER SUBJECTED TO A BATTLEFIELD THREAT



REFERENCES

1. B. D. Burns and S. Zuckerman, "The Relationship Between Striking Velocity and the Damage Caused by a 3/32 Inch Steel Ball," Great Britain, Ministry of Home Security, R. C. Report No. 232, 1941.

2. R. W. Guerney, "A New Casualty Criterion," Ballistic Research Laboratories Report No. 498, 31 October 1944.

3. T. E. Sterne, "A Provisional Casualty Criterion," Ballistic Research Laboratories Technical Note No. 370, March 1957.

4. F. Allen and J. Sperrazza, "New Casualty Criteria for Wounding By Fragments," Ballistic Research Laboratories Report No. 996, October 1956.

5. T. E. Sterne, "Provisional Criteria for Fatal or Severe Wounding By Fragments," Ballistic Research Laboratories Memorandum Report No. 591, February 1952.

6. R. Kneibert et al, "Wound Tract Analysis by High Speed Digital Computer," Ballistic Research Laboratories Report No. 1623, January 1965.

7. L. Sturdivan, J. Thompson, J. Terner, "A Parametric Study of Bio-Retardation," paper presented at Army Design of Experiments Conference, April 1969.

LIST OF ATTENDEES

Carmelo Aliano, Redstone Arsenal, Alabama
Bernard Alley, Redstone Arsenal, Alabama
Hubert W. Anderson, Redstone Arsenal, Alabama
Edward J. Arber, Picatinny Arsenal, New Jersey
Elwood D. Baas, White Sands Missile Range, New Mexico
Rolf Bargmann, University of Georgia, Athens, Georgia
Garry Barnard, Redstone Arsenal, Alabama
B. W. Barnett, Redstone Arsenal, Alabama
Edwin M. Bartee, Vanderbilt University, Nashville, Tennessee
Hans Baussus von Luetzow, Fort Belvoir, Virginia
Robert Bechhofer, Cornell University, Ithaca, New York
Alan W. Benton, Aberdeen Proving Ground, Maryland
Joe Bigger, Redstone Arsenal, Alabama
Novella S. Billions, Redstone Arsenal, Alabama
Erwin Biser, Fort Monmouth, New Jersey
Mary E. Blome, Ammunition Proc. and Supply Agency, Joliet, Illinois
Joseph I. Bluhm, Army Materials & Mech. Res Lab, Watertown, Mass
E. L. Bombara, NASA, Huntsville, Alabama
D. G. Boyd, Redstone Arsenal, Alabama
Helen M. Boyd, Redstone Arsenal, Alabama
Wallace Branham, Redstone Arsenal, Alabama
Ferrell Bryan, Redstone Arsenal, Alabama
O. P. Bruno, Aberdeen Proving Ground, Maryland
Edwin L. Buchanan, Redstone Arsenal, Alabama
Leo F. Buldhaupt, The Boeing Company, Huntsville, Alabama
Joseph M. Cameron, National Bureau of Standards, Washington, D. C.
D. Ray Campbell, Redstone Arsenal, Alabama
Vera M. Campbell, Redstone Arsenal, Alabama
Ola Carter, Redstone Arsenal, Alabama
Richard W. Clarke, Watervliet Arsenal, New York
A. C. Cohen, Jr., University of Georgia, Athens, Georgia
Charles Colvin, Redstone Arsenal, Alabama
Robert G. Conard, Redstone Arsenal, Alabama
John E. Condon, NASA Headquarters, Washington, D. C.
Gideon A. Culpepper, White Sands Missile Range, New Mexico
C. Alan Cummings, Redstone Arsenal, Alabama
Richard D'Accardi, Fort Monmouth, New Jersey
J. A. Davies, General Electric, Huntsville, Alabama
Carmine DeSanctus, NASA, Huntsville, Alabama
Henry A. Dihm, Jr., Redstone Arsenal, Alabama
James M. Donnini, The Boeing Company, Huntsville, Alabama
Francis G. Dressel, Army Research Office-Durham, North Carolina
Eugene Dutoit, Army Munitions Command, Dover, New Jersey
Alan R. Dyer, Aberdeen Proving Ground, Maryland
Churchill Eisenhart, National Bureau of Standards, Washington, D. C.
Oskar Essenwanger, Redstone Arsenal, Alabama
Wade H. Ewart, Redstone Arsenal, Alabama
Tom G. Feaster, The Boeing Company, Huntsville, Alabama

Preceding page blank

Walter T. Federer, University of Wisconsin MRC, Madison, Wisconsin
Lewis S. Fichter, Picatinny Arsenal, Dover, New Jersey
Thomas O. Finley, Fort Ord, California
Walter D. Foster, Fort Detrick, Maryland
Fred Frishman, Office, Chief of Research and Development, Washington, D.C.
A. S. Galbraith, Army Research Office-Durham, North Carolina
Sidney Gerard, Aberdeen Proving Ground, Maryland
John P. Gibbons, Redstone Arsenal, Alabama
Tom Gill, University of Alabama, Tuscaloosa, Alabama
Dieter Grau, NASA, Huntsville, Alabama
Bruce C. Gray, Fort Detrick, Maryland
Frank E. Grubbs, Aberdeen Proving Ground, Maryland
William V. Gudaitis, Redstone Arsenal, Alabama
John Gurland, MRC, University of Wisconsin, Madison, Wisconsin
Bernard S. Gurman, Fort Monmouth, New Jersey
William B. Harrell, NASA, Huntsville, Alabama
John S. Hagan, Aberdeen Proving Ground, Maryland
Bernard Harris, MRC, University of Wisconsin, Madison, Wisconsin
Boyd Harshbarger, Virginia Polytechnic Institute, Blacksburg, Virginia
Gary Hitchcock, Redstone Arsenal, Alabama
Earl G. Hoard, MASA, Huntsville, Alabama
Chester L. Kopkins, Redstone Arsenal, Alabama
John W. Howerton, Redstone Arsenal, Alabama
David R. Howes, STAG, Bethesda, Maryland
Edmund H. Inselmann, U. S. Army Materiel Command, Washington, D. C.
Jack Irwin, Redstone Arsenal, Alabama
Gerhard J. Isaac, Fitzsimons General Hospital, Denver, Colorado
Robert F. James, Redstone Arsenal, Alabama
William R. James, Coastal Engineering Research Center, Washington, D. C.
Toke Jayachandran, Fort Ord, California
Emil H. Jebe, Willow Run Laboratory, Willow Run, Michigan
Andrew H. Jenkins, Redstone Arsenal, Alabama
Jerome R. Johnson, Aberdeen Proving Ground, Maryland
Larry H. Johnson, Redstone Arsenal, Alabama
Wallace E. Jordan, NASA, Huntsville, Alabama
George L. Kinnett, Army Aviation Material Labs, Fort Eustis, Virginia
Raymond V. Knox, Redstone Arsenal, Alabama
Richard Krinsky, University of Alabama, University, Alabama
Richard G. Krutchkoff, Virginia Polytechnic Institute, Blacksburg, Virginia
Harry H. Ku, National Bureau of Standards, Washington, D. C.
William A. Lampe, Redstone Arsenal, Alabama
Raymond E. Lacy, Fort Monmouth, New Jersey
George I. Lavin, Aberdeen Proving Ground, Maryland
Patrick B. Lawler, Jr., Redstone Arsenal, Maryland
Siegfried H. Lehnigk, Redstone Arsenal, Alabama
George A. Lineberry, Redstone Arsenal, Alabama
Joseph Lovinger, Rohm and Haas Company, Huntsville, Alabama
H. L. Lucas, North Carolina State University, Raleigh, North Carolina
Bob B. Lukens, University of Alabama, Huntsville, Alabama

Mat V. Maddix, Redstone Arsenal, Alabama
Clifford J. Maloney, National Institute of Health, Bethesda, Maryland
Nancy R. Mann, North American Rockwell Corporation, Canoga Park, California
Nathan Mantel, National Cancer Institute, Bethesda, Maryland
Stanley M. Martin, National Communicable Disease Center, Atlanta, Georgia
John L. McDaniel, Redstone Arsenal, Alabama
Caludia K. McDonald, Redstone Arsenal, Alabama
William B. McIntosh, Aberdeen Proving Ground, Maryland
Ray L. Miller, Jr., Shaw Air Force Base, South Carolina
James N. Mitchell, Safeguard Logistics Command, Huntsville, Alabama
James R. Moore, Aberdeen Proving Ground, Maryland
Oskar Morgenstern, MATHEMATICA, Princeton, New Jersey
Jacques Naar, Stanford Research Institute, Menlo Park, California
William Nichols, Redstone Arsenal, Alabama
Henry Nocke, SPACO, Inc., Huntsville, Alabama
William H. Nuckols, III, Redstone Arsenal, Alabama
Richard L. Nutt, Redstone Arsenal, Alabama
Robert R. Orr, Redstone Arsenal, Alabama
Bernard Ostle, Florida Technological University, Orlando, Florida
Tsetsy Pangarova, Redstone Arsenal, Alabama
Albert Parent, Thiokol Chemical Corp., Huntsville, Alabama
Gene B. Parrish, Army Research Office-Durham, North Carolina
R. H. Peterson, Aberdeen Proving Ground, Maryland
Charles A. Pollard, Jr., Redstone Arsenal, Alabama
Nixon Powell, Redstone Arsenal, Alabama
Aubrey Presson, Redstone Arsenal, Alabama
Robert C. Raybold, National Bureau of Standards, Washington, D. C.
Vernon H. Rechmeyer, Thiokol Chemical Corp., Huntsville, Alabama
Nancy R. Rich, Redstone Arsenal, Alabama
B. A. Richardson, Canadian Forces Headquarters, Ottawa, Ontario, Canada
Carol D. Rose, Army Tank Automotive Command, Warren, Michigan
Don Royston, Thiokol Chemical Corp., Huntsville, Alabama
Doyce Satterfield, Redstone Arsenal, Alabama
Sam C. Saunders, Boeing Scientific Res. Lab., Seattle, Washington
Raymond B. Schnell, Army Materiel Command, Washington, D. C.
George Schroeter, Aberdeen Proving Ground, Maryland
Gordon Setterland, Boeing Company, Huntsville, Alabama
Harold C. Sebring, STANO Systems Office, Washington, D. C.
Seymour M. Selig, Department of the Navy, Washington, D. C.
R. E. Shannon, University of Alabama, Huntsville, Alabama
Chester E. Sharp, Fort Monmouth, New Jersey
Ralph Simenson, Redstone Arsenal, Alabama
Leslie E. Simon, USA Ret, Winter Park, Florida
Orvel E. Smith, NASA, Huntsville, Alabama
Royce W. Soanes, Watervliet Arsenal, New York
Earl B. Stewart, Redstone Arsenal, Alabama
R. G. Stimson, Department of the Army, Washington, D. C.
Glenn A. Stoops, Fort Ord, California
James F. Strickland, Jr., Aberdeen Proving Ground, Maryland

Donald H. Strietzel, Safeguard System Command, Huntsville, Alabama
William C. Taylor, Aberdeen Proving Ground, Maryland
Jerry Thomas, Aberdeen Proving Ground, Maryland
R. M. Thrall, Rice University, Houston, Texas
Myrna Toivanen, Honeywell, Inc., Minneapolis, Minn.
Joseph Weinstein, Fort Monmouth, New Jersey
Burchard E. Wheatley, The Boeing Company, Huntsville, Alabama
Robert D. Williams, Stanford Research Institute, Menlo Park, California
Rosalie C. Woodall, Harry Diamond Laboratories, Washington, D. C.
W. J. Youden, National Bureau of Standards, Washington, D. C.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

ORIGINATING ACTIVITY (Corporate author) U. S. Army Research Office-Durham Box CM, Duke Station Durham, North Carolina 27706	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A
REPORT TITLE PROCEEDINGS OF THE FIFTEENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Interim Technical Report	
5. AUTHOR(S) (First name, middle initial, last name)	
6. REPORT DATE July 1970	7a. TOTAL NO. OF PAGES 882 7b. NO. OF REFS
8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ARO-D Report 70-2 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
10. DISTRIBUTION STATEMENT This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of the U. S. Army Research Office - Durham.	
11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Army Mathematics Steering Committee on behalf of the Office of the Chief of Research and Development

13. ABSTRACT

~~This is the~~ technical report resulting from the Fifteenth Conference on the Design of Experiments in Army Research, Development and Testing. It contains 35 papers presented at that meeting. These treat various Army statistical and design problems.

14. KEY WORDS:

reliability
performance capability
estimation of variance
quality control
statistical models
Multiple triangulation
interlaboratory studies
Bayesian confidence bounds
real-time simulation
continuous sampling
weapons effectiveness
air defense systems
manpower planning
disconnected block designs
design of field tests
infrasonic waves
least squares
assay procedures
linear programming
catastrophic threat
empirical Bayes methods
man-machine performance
orthogonal latin squares
confidence limits

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

unclassified

Security Classification