THE UNIVERSITY OF MICHIGAN
INDUSTRY PROGRAM OF THE COLLEGE OF ENGINEERING

Planning a Decisive Experiment

Transcript of the Fourth Ann Arbor Industry-Education Symposium

December, 1959
IP-410

TABLE OF CONTENTS

	Page
PROGRAM COMMITTEE MEMBERS	iii
INTRODUCTION	iv
OPENING REMARKS S. S. Attwood	1
THE THEORY OF DECISIVE EXPERIMENTS Paul Henle	3
INTEGRAL VERSUS DIFFERENTIAL EXPERIMENTS S. W. Churchill	13
SCREENING SOLUTIONS TO UPGRADE QUALITY: A NEW APPROACH TO PROBLEM SOLVING UNDER CONDITIONS OF UNCERTAINTY Norman R. F. Maier	27
APPROACH TO A PROBLEM IN MICROBIOLOGICAL CONVERSION Durey H. Peterson	47
SOME TECHNIQUES FOR OPTIMIZING AN EXPERIMENTAL DESIGN R. B. Morrison	59
THE USE OF MATHEMATICAL MODELS IN PLANNING AN EXPERIMENTAL PROGRAM Wilson P. Tanner, Jr.	75
GOOD AND BAD RESEARCH Peter Franken	85

PROGRAM COMMITTEE MEMBERS

J. W. Greiner
The Upjohn Company

Harry H. Holscher
Owens-Illinois Technical Center

Fred L. Black
Industry Program

Raymond E. Carroll
Industry Program

Glenn V. Edmonson
College of Engineering

William Kerr
Nuclear and Electrical Engineering

Richard B. Morrison
Aeronautical Engineering

Robert R. White
Chemical and Metallurgical Engineering

Brymer Williams
Chemical and Metallurgical Engineering

J. Louis York
Chemical and Metallurgical Engineering

INTRODUCTION

The symposium committee tried a number of "organizing frameworks" for this program, so that the presentations could cover major considerations in an orderly fashion. We realized rather soon that the subject was much too broad and diverse for this treatment; a completely different approach developed.

Knowing that many basic principles of the experimental method are common to all research, the committee decided to invite seven experts from seven different fields, leaving each man free to talk about decisive experiments from his own particular viewpoint and experience. The resulting program presented many specific ideas as well as general methods of approach which, although they were described in terms of one discipline, can well be incorporated in almost any other field of research.

The talks were recorded on tape. The transcriptions were edited as little as possible in most cases, to preserve the oral presentation style of the speakers.

This record is not a publication, and has been reproduced primarily for the benefit of those who attended the meeting.

OPENING REMARKS

S. S. Attwood

Dean of the College of Engineering

I want to welcome you all here to the Fourth Annual Industry Education Symposium. As the years go on and we repeat these experiences, it is with increasing pleasure and satisfaction that we have a day in which the representatives from the University and representatives from industry can get together to discuss their mutual interests. I understand that while probably most of you come from Michigan or a nearby area, some have come from considerably greater distances away. But you are all welcome no matter what the distance is that you may have traveled.

When Ray Carroll first suggested this program, "Planning a Decisive Experiment," I could instinctively feel that it was an important type of subject. But I must say at the time I was unable to figure out just how he was going to work it out. But as I look over the program which has finally emerged, I think he and the committee have done a marvelous job in bringing together a considerable number of people, all of whom can contribute significantly to answering this question that has been raised. And then I think, too, that you will enjoy the windup of the afternoon when we will have our informal discussion with various members of the faculty and other gentlemen from industry. So I look forward quite enthusiastically to listening to these papers myself and I am sure that you will also.

I would like to revert to my experience yesterday. The term "decisive experiment" rings a bell this morning. Yesterday I had the privilege of being invited up to Bay City to witness the launching of a new boat at the Defoe Shipyards. It's a naval vessel...many of you may well have read about it in the paper this morning. It's a destroyer of a new kind, named the U.S.S. Henry B. Wilson, after Admiral Wilson who was a long-time naval officer and served as Superintendent of the Naval Academy. It was sponsored by his daughter who is now Mrs. Patrick J. Hurley. Mrs. Hurley and General Hurley were both there yesterday and helped make the occasion very successful. General Hurley, you will remember, was Secretary of War during part of the Hoover administration. Why I call it a "decisive experiment" up there was that in the first place it was the first boat of that size built on the lakes for sea duty. Until the opening of the St. Lawrence Seaway, they couldn't possibly have gotten such a boat through. It is some 450 feet long, and I have forgotten what the

beam is, and its' draft is some 23 feet. It is composed of a very special steel; the type that they use in submarines which is very much tougher and stronger than ordinarily used in boats. But here it was, sitting up there, an enormous looking thing when you stood beside it (and some 8,000 of us did), and the question in our minds was "will it slide in sideways just right and flop in the water the way it should?" It certainly seemed to all of us present that this was very much of a decisive experiment. What could have gone wrong? I could see the men there who had charge of the work very much concerned as they might well be, but actually the six ropes which held it were cut by guillotines which were electrically operated so they all chopped together and in five seconds it was in. We all heaved a great sigh of relief and then we went off to the Winona Hotel. An admirable lunch, along with cocktails, was served to us. So! I think that you will all enjoy our program today concerning the decisive experiment. We are very happy to have you here.

THE THEORY OF DECISIVE EXPERIMENTS

Paul Henle

Professor of Philosophy
The University of Michigan

THE THEORY OF DECISIVE EXPERIMENTS

Paul Henle

What I would like to do in this discussion is to give some indication of the theory of experiments ... I won't quite say the theory of decisive experiments because I am afraid I will have to conclude that there aren't any completely decisive experiments (I hope Mr. Carroll doesn't mind...) but, to indicate very briefly and necessarily sketchily something of a model for the way in which any experiment would work - something of an indication of the logic of experiments. Now in doing this, what I shall try to do is very briefly present what I take to be a model of the situation. This will be only one interpretation of the way in which experiments work, but I shall try to indicate that it is at least a plausible one. And in a sense what I will be going over is material that you already know. After all, you are acquainted with this sort of process. The only thing that I shall hope to do is to make it a little bit more articulate, and a little bit more explicit than it was before.

As I indicated, I want to try a probability model for this situation. Let's start with a very simple problem of probabilities and since it's traditional that probability problems concern themselves with beans and bags, I don't see why we should depart from that custom. Suppose then, we have two bags. One of them contains nothing but black beans. The other contains half black beans and half white beans. And suppose that a person is blindfolded, draws from one of these bags and gets a black bean. What is the probability we may ask that he was drawing from the bag which contained nothing but black beans. Well, if we want to work with this illustration at all, we need a little bit of terminology, and seeing that the terminology of probability differs from author to author, I am going to make this explicit. Let us write

$$P(bA)$$

meaning the probability of drawing a black bean from bag A. Bag A will be the one which contains all black beans. Then we can say, keeping this terminology, the probability of getting a black bean from the other bag can be written this way, with the bar over the A to indicate not-A

$$P(b\bar{A})$$

Now we can assign values to these. The probability of drawing a black bean from bag A is 1 (they are all black). The probability of getting a black bean from the other bag is 1/2 (half of them are black). But

now what we are interested in is not this very simple question, but rather the probability that we drew from bag A, given we drew a black bean, and although we are not interested in it, we can also raise the question, what is the probability that we were drawing from the other bag, given that we drew a black bean. With this much notation we can go ahead to solve our problem. There is a very simple rule of probability, easily proved....I would prove it for you if it weren't quite so early in the morning and I'm not very good at proofs early in the morning..which says that the ratio of these probabilities

$$\frac{P(bA)}{P(b\bar{A})}$$

is equal to the ratio of these probabilities

$$\frac{P(Ab)}{P(\bar{A}b)}$$

That is, if we take the probability of getting a black bean from bag A, and divide it by the probability of getting a black bean from the other bag, this will give us the probability that we were drawing from bag A after we have drawn the black bean, divided by the probability that we are drawing from the other bag. This ratio of probabilities can be regarded as odds. That is, the odds on one thing against its not happening is the probability that it will happen divided by the probability that it won't. So what we have here is the odds on drawing a black bean if we are drawing from bag A or if we are drawing from bag $\bar{A}$, and this in turn gives us the odds that we were drawing from bag A as against the bag which isn't A, which has the mixture of beans. So, seeing that we have numerical values here, the probability of $P(bA)$ was I, the probability of $P(b\bar{A})$ was $1/2$, the value of this ratio is 2 to 1, we can say that:

$$\frac{P(bA)}{P(b\bar{A})} = \frac{1}{1/2} = \frac{2}{1} = \frac{P(Ab)}{P(\bar{A}b)}$$

In other words, the odds are 2 to 1 that this bean came from the bag which contains all black beans. Well, this is a very simple sort of problem and by itself not worth much. But let's complicate it a little bit. Suppose that instead of having just one bag, of type A with all black beans, we have several of them. Let us say four to be exact. And let us say that of the other sort of bag, which is half white and half black, we have three of them. This obviously makes a difference because seeing that there are more bags which contain all black, we increase the probability of having drawn from one of them and we may then raise the question: How does it influence the probability? And in order to see that, we have to introduce one other set

of odds, one other ratio of probabilities. Let us indicate by $P(A)$, the probability before any draw is made that we will draw from a bag of beans, all of which are black, and by $P(\bar{A})$ the probability, given the circumstances that we will draw from a bag which contained partly white and partly black beans. With the setup that I have given you, with four of the type A and three of type $\bar{A}$, the probability, of course, of drawing from a bag of type A would be $\frac{4}{7}$, the probability of the other would be $\frac{3}{7}$.

Once again, having brought these probabilities into consideration, one can show that they must be used in our formula and we must, in order to figure the odds that we were drawing from an A bag consider not merely the ratio

$$\frac{P(bA)}{P(b\bar{A})}$$

but we must also allow for the antecedent probabilities of these two situations and must include

$$\frac{P(A)}{P(\bar{A})}$$

These ratios may be combined into the formula

$$\frac{P(Ab)}{P(\bar{A}b)} = \frac{P(bA)}{P(b\bar{A})} \times \frac{P(A)}{P(\bar{A})}$$

Now once again, I am not going to stop for the derivation of this equation, but it follows very easily from simple rules of probability. It says that the odds that we drew from a bag of type $\underline{A}$ is a product of two sets of odds -- the odds that we would draw a black bean if we were drawing from a bag of type $\underline{A}$ and the antecedent odds on drawing from a bag of type $\underline{A}$. This is a perfectly straightforward problem in probabilities and I feel a little bit ashamed at bothering you with it. But now let's change the conditions still a bit more. I have given numerical values to the probability

$$\frac{P(bA)}{P(b\bar{A})}$$

and to

$$\frac{P(A)}{P(\bar{A})}$$

Let's relax these conditions a little bit. Suppose we don't quite know the probability of drawing a black bean from bag A, we are not

sure of its composition. All we know is that, let us say, at least 90% of the beans in bag A are black. We can't say definitely that the probability is 9/10, but we could say that it is at least 9/10ths. Let us say here that we don't know how many black and how many white beans there are in the rest of the bags. They may differ among themselves, but at least they contain no more than 1/4 of black. In these terms we can't fix a definite ratio for fraction

$$\frac{P(bA)}{P(b\bar{A})}$$

but we can at least fix limits within which it will vary, and in this way we can still see that the ratio that we get here will add something to the probability.

Again, let's vary this situation with regard to antecedent probabilities. We don't know exactly how many bags there are of each type, but we know there is something like the same number, although we are not quite sure, so roughly this factor could be considered something like one. Or, to change it, perhaps this factor could be considered only something like 1/10. Or perhaps we are not even sure of that, but we think it's less than 1/2, although not much less than 1/2. Still, with all these limitations, we can operate with our formula and we can find out something about how much the drawing of the black bean will add to the initial probability. We will not be able to state this with any numerical exactitude, but we can still have an increasing confidence in our theory in a way which we couldn't have had at the beginning.

This is what I wanted to get out of the model. This is all that we need this model for, and as we have put the thing now, instead of considering merely beans in a bag, let's consider any sort of hypothesis whatsoever. Here I think one can say that the same sort of conditions apply. Let us take as a sort of stock example the old theory that fluctuations in the stock market vary with sunspots. There has been some evidence to this effect, but I'm not here to try to convince you that that is, or is not the case. Let's just consider it as a sample. Let $\underline{A}$ be the hypothesis that the activity of the stock market varies with the activity of sunspots. How you want to measure the activity doesn't much matter. Let's assume that it's the volume of sales that we are taking into account, so that whenever you have a lot of sunspots, you have a heavy volume - when the sunspots slack off, so does the stock market. What is the probability of $\underline{A}$ to start with -- well, quite small. How small, I don't know, but it doesn't matter. What is the probability of $\underline{\bar{A}}$, $\bar{\underline{A}}$ here would be the probability that the activity of the stock market doesn't depend on sunspots. And the value for this fraction would be very large. Well and good. Let's suggest

some sort of investigation. Suppose we find a given point of sunspot activity, find it to be quite high -- we investigate the market on that date and find that the volume of sales was very large. We will call this whole investigation, this finding of ours, b. What is the probability that B would occur -- that is, that there would be a coincidence of a heavy market and sunspots, on the basis of the theory that the two vary together? That probability is going to be 1, or very close to 1. What is the probability that you would get a heavy market if the sunspots were not the deciding factor in the market? What you would do to get an answer to that is see how often the market is heavy, figure out what percentage of the time, and then give that as the ratio here. So, one could get, roughly at least, some sort of ratio here. And what one would have then, is a fraction whose numerator is 1 and whose denominator is less than 1, an improper fraction. On the basis of that, we would say that the odds on A, given over here had been increased by this particular situation, so that however small the odds on A were to begin with, the odds on A after this investigation emerge as greater than they were before. And in this way we can say that the way in which this investigation confirms the theory is like our probability situation, and we can use the probability situation as a model. This, if you like, is the theory of the way in which some sort of investigation confirms a hypothesis. But what would it mean to make the experiment decisive, so that we would decide once and for all that whatever the odds on A had been at the beginning, after this experiment, they would have to be overwhelmingly in favor of A after it. Obviously this depends on the value of the fraction

$$\frac{P(b,A)}{P(b\,\overline{A})}$$

and the larger we can make this value, the nearer we come to a decisive experiment. If we could get a situation where one certain event, or certain outcome of an experiment, whose probability was 1 on A and 0 on $\overline{A}$, then this thing would be infinitely large, and we would have a decisive experiment, and this seems to be the only way in which one could get something which was absolutely decisive. Well, how do you design an experiment the outcome of which will be certain on one hypothesis and impossible on another? This, so far as I know, cannot be done. There was one famous example in the 19th Century where they thought they had done this, but even then it turned out that it hadn't happened. You probably recall the situation. There had been competing theories as to the nature of light. According to one point of view, light was a wave motion. According to another point of view, it was a motion of corpuscles. There were differences on what you could predict on the basis of this, but the differences were not at first testable. You could predict that if light went into a heavier medium, then according to the corpuscular theory, it should move faster. According to the

wave theory, it should move slower. But up to the middle of the 19th Century, no one had an apparatus which could measure the speed of light accurately enough. Finally, in the 1850's, Foucault did it. He couldn't determine the speed of light accurately, but at least he could show it was less in water than it was in air. This looked like the decisive experiment. This looked as if you had a probability of 1 of an occurrence here, of 0 in the other place, and this is in the textbooks as one of the very few decisive experiments in the history of science. But, corpuscular theories, or something like them are going today and it was found that when the situation was analyzed, what was being tested was not merely a corpuscular theory, but a corpuscular theory with certain other assumptions -- assumptions as to the attraction of particles involved, and it turned out that it wasn't that a corpuscular theory by itself made Foucault's result impossible, but merely some of the additional assumptions that went with it. And with that disappointment, I think that what one has to say is that any hope of devising an iron-clad, decisive experiment which would settle the question once and for all is gone. But on the other hand, decisive experiments aren't necessary. What one wants is not to get the problem settled once and for all. This would be ideal, of course, but this isn't the way in which things work. All one wants is a reasonably decisive experiment and if in terms of the model, we ask, "How do we get a reasonably decisive experiment?" The answer is quite simple. Make the fraction

$$\frac{P(b,A)}{P(b\,\bar{A})}$$

as large as possible. Make a prediction which would certainly be the case on the hypothesis being tested, and it would be very unlikely to be the case if the hypothesis were false. And the more unlikely it would be to be true apart from the hypothesis, the more decisive your experiment would be.

One conclusion then which I draw from this very simple model is: you don't get a decisive experiment, but your experiment becomes more decisive as this ratio becomes larger. Well, the next question might be, "This is all very well, but how do you know when your experiment is decisive enough?" And here, it seems to me, there isn't any clear-cut answer. In part it depends on the expectancies. How much depends on this experiment being correct. If you don't have to bet too heavily, you are more likely to be satisfied with less evidence. The other point would be that quite apart from this, you can get certain measures of how decisive an experiment is. In the social sciences they very often rate answers according to how likely this result would be to occur if the phenomenon were to be explained merely in chance and if a person gets a result so there is only one chance in 100, he would get his result by chance rather than by the correctness

of his theory, then he feels quite good. If he finds that it's only one chance in 20, very often he's reasonably satisfied. But still if you take one chance in twenty, or one chance in one hundred, so far as I can see, there is no answer to this question. No general answer and it is for this reason that there are experts in the field. That is, it seems to me that here is a case where one can't depend on any rule, but must use some sort of judgment. If it weren't for that, we would have what for me would be a delightful situation that a logician could take over any sort of situation, but just because not everything works according to rule, one has to know something about the field in which he is working, and there is a legitimate place for experts.

INTEGRAL VERSUS DIFFERENTIAL EXPERIMENTS

S. W. Churchill

Professor of Chemical Engineering
The University of Michigan

INTEGRAL VERSUS DIFFERENTIAL EXPERIMENTS

S. W. Churchill

One function of engineers is to predict chemical and physical behavior. The experimental data we obtain in the laboratory provide the basis for these predictions. We collect, summarize and generalize the data in the form of correlations. As a basis for correlation we often use models. The role of theory in engineering is primarily to provide us with models for correlation.

It is becoming the practice in the technical literature to present only the model and the author's conclusions and not the data. We may blame the editor for this practice but it is apparently accepted by the majority of readers and authors. Many authors appear to be prouder of their conclusions and models than of their data. However, if we do not have access to the original data we cannot evaluate the validity of the author's conclusions. Rationalization without reference to the quality and quantity of the data is meaningless speculation.

This morning we will examine the problem of obtaining and interpreting data for some typical rate processes. In general the rate is defined as the change in some material per unit time per unit extent of the system. The size of the equipment necessary to produce a given change is inversely proportional to the rate at which the process is carried out. Therefore the rate is a very important quantity in process design.

As a first example let us consider a homogeneous chemical reaction. The process can be characterized by the number of moles of some reacting species A present in the reactor at any time and the rate of reaction can then be defined as the decrease in the number of moles of A per unit time per unit volume, or mathematically as

$$r_A = - \frac{dN_A}{V d\theta} \qquad (1)$$

where θ = time
N_A = moles of A
V = volume of reacting system

The definition of the rate provides us with a set of instructions for measurement of the rate. Equation (1) indicates that we need to know the number of moles of A and the volume of the reacting system as a function of time. Measurements of composition, temperature and pressure versus time permit us to calculate the number of moles of A as a function of time from material balances and an equation of state. These calculations are usually minimized by using a reactor of constant volume and maintaining the temperature constant and uniform. The calculated number of moles of A must then be differentiated with respect to time and the derivative divided by the volume of the reacting system.

This differentiation might be accomplished by plotting N_A versus θ, drawing a curve through the data points and then reading the slope of the curve as a function of θ or N_A. This procedure is not very satisfactory. The location of the curve is highly arbitrary with most sets of real data and the slope depends critically upon the type of curve drawn through the data. The author's prejudices are unavoidably embodied in the data at this point by the type and location of the curve that he draws. In addition the values determined for the derivative give no indication of the uncertainty and experimental scatter in the raw data.

A better procedure is as follows. The average rate over each interval of time between the measurements is calculated as indicated below:

$$(r_A)_{ave} = -\frac{N_{A2} - N_{A1}}{V[\theta_2 - \theta_1]} = -\frac{\Delta N_A}{V\Delta\theta} \qquad (2)$$

Now what we want is the instantaneous rate, i.e., the limit of the average rate as the time interval goes to zero:

$$r = \lim_{\Delta\theta \to 0} \left(-\frac{\Delta N_A}{V\Delta\theta} \right) = \frac{dN_A}{Vd\theta} \qquad (3)$$

How can we carry out this process in practice? If we reduce the time interval between measurements we soon have changes in composition that are less than the uncertainty in our chemical analyses. One way of approximating the limiting process is to plot the data as $-\Delta N_A/\Delta\theta$ versus θ as illustrated in Figure 1. The average rates are plotted as steps over the corresponding intervals of time. A curve having the same area below as the steps provides an estimate

of the instantaneous rate since

$$\text{Area under steps} = \left(\frac{\Delta N_A}{\Delta\theta}\right)\Delta\theta = \Delta N_A = \int_{N_A}^{N_A+\Delta N_A} dN_A = \int_{\theta}^{\theta+\Delta\theta}\left(\frac{dN_A}{d\theta}\right)d\theta = \text{area under curve} \quad (4)$$

This is a mathematical proof, if you like, that the instantaneous rate should be an equal area curve through the steps representing the average rates. Two difficulties arise. Firstly, there are an infinite number of equal area curves which could be drawn through the steps. If we didn't know anything about the process we might even draw an oscillating curve. Our experience with other chemical reactions suggests however that a simple monotonic curve is more probable. Secondly, no one ever obtains as good data as those in Figure 1. We will subsequently examine some real data. The prejudice introduced in the arbitrary construction of the equal area curve is much less serious than in the construction of an arbitrary curve through the plot of N_A versus θ since in the finite difference plot we can see the uncertainty in the average values of the rate directly.

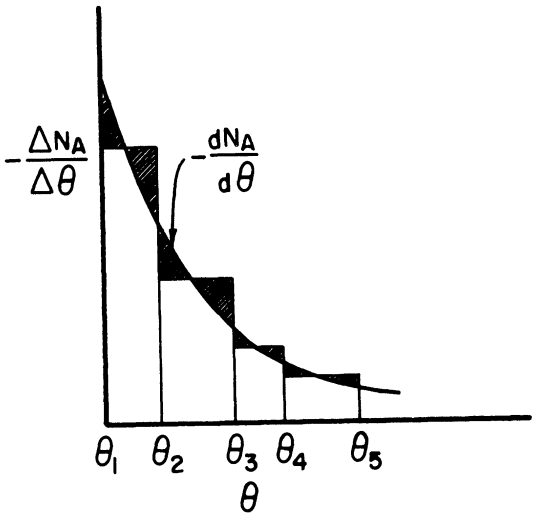


Figure 1.

After drawing the equal area curve we could read values of the instantaneous rate at the measured values of time and attempt to correlate the rate with the variables such as temperature, pressure and composition that describe the environment in which the reaction took place. The correlation process is one of trial and error. Models are postulated and tested until a satisfactory representation for the rate is obtained. The constants in the model are usually evaluated by some standard procedure such as least squares even though the number of data points is insufficient to give formal support to such statistical procedure. Theory enters for the first time at this stage in that it may suggest one or more models for correlation.

The definition of the rate also provides us with a set of instructions for process design. The time required to accomplish a chemical conversion is given by the integral obtained by rearrangement of Equation (1):

$$\theta = \frac{1}{V} \int_{N_{A1}}^{N_{A2}} \frac{dN_A}{r_A} \qquad (5)$$

In order to carry out this integration it is necessary to know the rate as a function of the number of moles of A in the system. A correlation for the rate as a function of composition, temperature and pressure together with material and energy balances and an equation of state provides us with the functional relationship. The integral may be carried out analytically, numerically or graphically depending upon the form of the rate function, etc.

We can use this design integral to process rate data. First we postulate a mathematical model containing arbitrary constants for the reaction rate and carry out the integration in Equation (5) to yield a relationship between time and the number of moles of A in the system with the constants as parameters. Then we use the experimental values for composition and time to evaluate the constants in this integral relationship. The entire procedure may be repeated with other models until a satisfactory representation for the data is obtained. We will compare this integral procedure with the differential procedure.

The data to be examined were obtained by passing a gas stream continuously through a granulated bed of catalyst in a long tube. The reaction was observed to be of the type:

$$A \leftrightharpoons R + S \qquad (6)$$

The outlet composition was measured for a series of flow rates and pressures. The rate of reaction in this type of reactor can be defined as the decrease in the number of moles of some species A per unit time per unit mass of catalyst. If the rate of flow of species A into any differential volume of the reactor, dV, containing a differential mass of catalyst, dW, is n_A moles per unit time, and the rate of flow of A out of the differential volume is $n_A + dn_A$ the rate can be expressed mathematically as:

$$r_A = - \frac{dn_A}{dW} \qquad (7)$$

This expression for the rate can be expanded in terms of the mole fraction of A, x_A, and the total molal flow rate, n_T; or in terms

of the molal rate at which A enters the reactor, n_{AO}, and the fractional conversion of A, Z_A, as follows:

$$r_A = - \frac{dn_T x_A}{dW} = + n_{AO} \frac{dZ_A}{dW} = \frac{dZ_A}{d(W/n_{AO})} \qquad (8)$$

The definition of the rate indicates that the rate can be determined from measurements of conversion as a function of the feed rate with a fixed mass of catalyst. The corresponding process integral is,

$$\frac{W}{n_{AO}} = \int_0^{Z_A} \frac{dZ_A}{r_A} \qquad (9)$$

The experimenter postulated a number of models, carried out the integration indicated in Equation (9), and evaluated the constants in the models by least squares using his experimental data for the conversion of A as a function of the catalyst-to-feed ratio W/n_{AO}. The model which he decided best fit the data was:

$$r_A = \frac{k(p_A - \frac{p_R p_S}{K})}{1 + K_A p_A + K_R p_R + K_S p_S} \qquad (10)$$

where p_A, p_R, and p_S are the partial pressures of A, R, and S. This particular model contains five constants k, K, K_A, K_R, and K_S. In Figure 2 the experimental data are compared with the model. The curves represent the model with the empirically determined constants. Total pressure was a parameter in the investigation and the different curves and different symbols represent different total pressures. We can see that there is moderate agreement between the data and the model. Those of you who have experience with catalytic reactions will recognize that these are reasonably good data for such a system.

Figure 2 tells us how the conversion varies with the catalyst-to-feed ratio and indicates the scatter of the data. What we really want to know is the rate and how well the model predicts the rate. We can use Equation (10) with the empirically evaluated constants to predict the rate but we would not have a direct indication of the reliability of the predictions of rate.

Let us examine the differential method of determining the rate from these data. In Figure 3 the average rates calculated from the data are plotted versus the catalyst-to-feed ratio as suggested by Equation (8). Equal area curves have been sketched through the steps representing the average rates. When we look at the data in this form they lose their apparent precision. We have difficulty in seeing any trend in the rate with the catalyst-to-feed ratio or with the total pressure.

What has happened to the apparent precision of the data? Differentiation has increased their uncertainty by about one order of magnitude. Why differentiate then? Because the rate and the uncertainty in the rate is what we want to know. The size of the equipment we design for the process will be inversely proportional to the rate and the safety factor we must include in the design will be a function of the uncertainty in the rate.

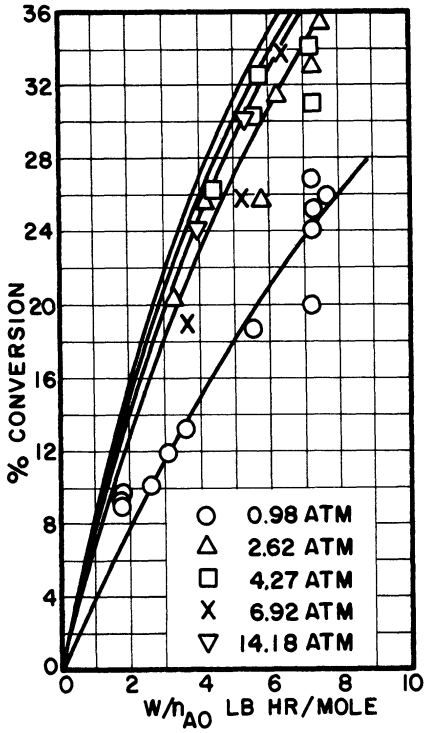


Figure 2.

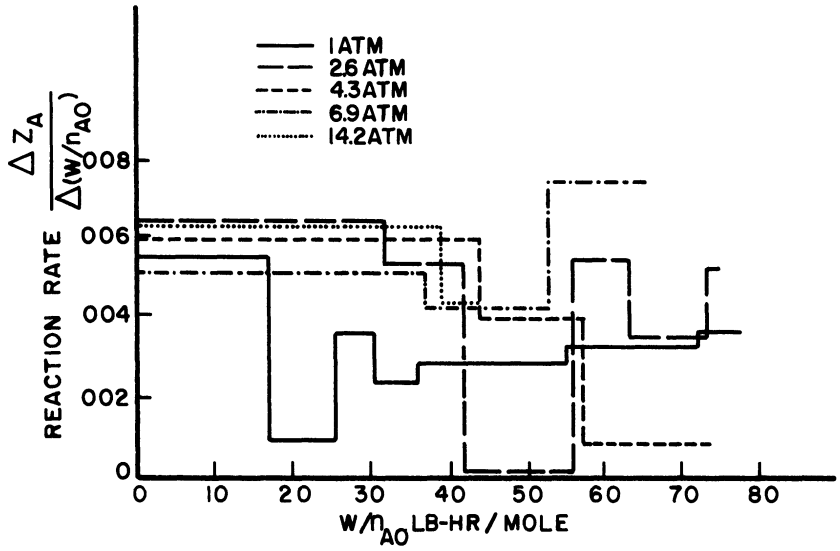


Figure 3.

Why don't we see plots of this type in the literature more often? For obvious reasons the experimentalist who works in the laboratory a year or more to obtain the data does not want to put his data in this form which shows it in the worst possible light. He is quite content to leave it in the form of Figure 2. He may even prefer to leave off the data points in Figure 2.

The designer, however, wants to know the worst. He must know how big a safety factor to use. He wants to know the rate as precisely as possible but he is even more interested in knowing what the possible limits of the rate are. He prefers to see the data in the form of Figure 3.

If we look back at Figure 2 now, we can see the uncertainty in the rate, that is, the uncertainty in the slope of the relationship defined by the data points. This uncertainty is disguised in Figure 2 however, and the curves give the illusion of a better relationship than would the points alone. In retrospect we see that straight parallel lines could have been drawn through the points in Figure 2 without doing violence to the data. Such straight lines would correspond to a set of parallel horizontal lines in Figure 3, that is to no variation of rate with catalyst to feed ratio.

Of what value are these data? I assert again that they are not bad data. They are probably better than the average of the data used to design existing reactors and most of these reactors are performing admirably today. We can extract quite a bit of rate information from Figure 3. The rate varied only between say 0.02 and 0.08 moles per hour per pound of catalyst under the experimental conditions. If the plant conditions are to be the same we can use a mean value of say 0.06 without missing the design of the reactor by more than a factor of 2 or 3 and most design engineers would say that this is acceptable today.

If we were going to develop a correlation for the rate in terms of composition, pressure and temperature we would hardly expect to use a very complicated equation. A rough rule of thumb is that the data should stand one differentiation without becoming random, for each constant in the correlation. Can you imagine what the second or third derivative of these data would look like? We now wonder why the author had sufficient confidence to use a five constant equation for correlation. He even went one step further, and I quote from his articles: "It has been shown that a single-site surface reaction is the controlling mechanism....." This statement is of course pure

nonsense. We can represent the data with his five constant equation or with any five constant equation for that matter, since the data do not support more than one or two constants. It is all right to use any equation we please to represent the data but it is not all right to imply that we know something about the energy of activation, the absorption equilibrium constant and the mechanism of the reaction. That is, we are not justified in placing a physical interpretation upon these excessive constants.

The author was mislead because he examined his data in integral form. If we are interested in the rate we should examine the data in the form indicated by the definition of the rate. If we are interested in the uncertainty in the rate, and I assert that we always should be, we should preserve the uncertainty in the raw data during the process of correlation. If we put a curve through the data and then use values from the curve for correlation we have no measure of the validity of the correlation.

Our ability to construct models, through the assistance of theory, and our ability to test and process data, through the assistance of electronic computers, has far out run our ability to make rate measurements. Every time we wish to add a constant to a correlation or model we need to improve the precision of the experimental measurements by an order of magnitude. As you all know this is very difficult to do. We badly need a break-through in instrumentation.

This problem is not restricted to catalytic chemical reactions. It is true in heat transfer. It is certainly true in mass transfer. It is true in every rate process that I have had the opportunity to examine, although it is difficult to discover because, as I said, the raw data are rapidly disappearing from the technical literature.

I will show you a few more brief illustrations of the difference between integral and differential tests of experimental data. In 1908 King obtained data for the rate of heat transfer from a cylinder to a stream of gas. He heated a wire electrically, measured the current and voltage, and computed the temperature of the wire and the rate of heat transfer. He then developed a theoretical model for the rate of heat transfer based upon ideal fluid flow around the cylinder. This model gave him the local rate around the circumference of the cylinder as shown by the solid curve in Figure 4. This is a polar diagram--the radial distance from the cylinder to the curve is proportional to the local rate of heat transfer. King integrated around the cylinder and obtained an expression for the overall rate of heat transfer. This expression agreed reasonably well with his overall data. In fact people have been taking additional data for this system for

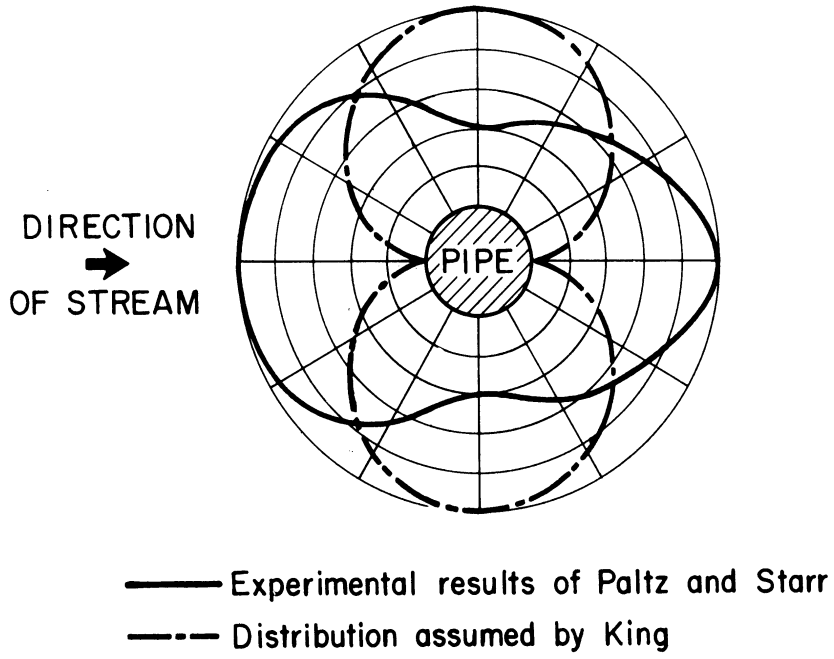


Figure 4

50 some years and his equation predicts these data within 10 or 20 percent. This agreement led people to trust the prediction of his equation for the local rate. About thirty years later Paltz and Starr measured the local rate of heat transfer around a cylinder. The dashed curve in Figure 4 represents these data. King's equation predicts a maximum rate at the sides of the cylinder and a zero rate at the front and back. His model couldn't have been more wrong; the maximum rate is in front, a lesser maximum occurs in back and the minimum rate is at the sides. Integration has the opposite effect of differentiation on data: all disturbances are smeared out and the uncertainty is reduced by an order of magnitude. Consequently the integral of King's equation could satisfactorily fit the integral data even though it failed for the local or differential data. This is also an illustration that overall or integral data are not a very sensitive test of a differential model.

The heat transfer coefficient for a fluid flowing through a pipe can be computed from the following dimensionless equation:

$$\frac{hD}{k} = \frac{2}{\int_0^1 \frac{u}{u_m}\left[\int_{r/r_w}^1 \frac{q}{q_w}\cdot\frac{k}{k_e}\, d\left(\frac{r}{r_w}\right)\right] d\left(\frac{r}{r_w}\right)^2} \qquad (11)$$

We won't bother with the details of this equation except to note that it involves two integrations and that we need to know the effective thermal conductivity, k_e, the velocity, u, and the radial heat flux, q, as functions of radius. Ten or so theoretical equations have been derived for heat transfer in pipes by assuming different expressions for the velocity, effective thermal conductivity and radial heat flux and carrying out the integrations. They all fit the data very well. I remember someone saying at a technical meeting a few years ago "It is rather remarkable that no matter what one assumes for the velocity distribution, one still obtains a good estimate of the rate of heat transfer." The reason of course is the double integration. The agreement does not mean that the assumed models for the velocity, effective thermal conductivity, and radial heat flux are correct nor does it indicate that these models will predict accurate values of the local rate of heat transfer.

The equations used to describe momentum transfer in a fluid are called the Navier-Stokes equations. They are three nonlinear, coupled, partial differential equations and are virtually insoluble except for highly idealized circumstances. If we could solve these equations we think we could predict the behavior of most fluid, mechanical systems. The equations are based on an analogy between the plastic deformation of a solid and the deformation of a fluid under shear and normal stresses. What would constitute a test of this model? If we could combine the three equations into one equation we would have at least fourth order derivatives of the velocity. Therefore we need velocity distribution data that are sufficiently precise to stand four differentiations. There are no data in the literature that even vaguely approach this quality. The Navier-Stokes equations are today a more complex model than we can use and we have no way of saying whether they are correct or not. Now people frequently say that they have confirmed the Navier-Stokes equations but what they have done is to find some idealized system where most of the terms in the equations drop out, integrate and then compare the integral form with integral data. As we have seen this does not constitute any test at all of the generality of the equation under other circumstances.

High-Speed Computation

Major influence on the rate processes today is the increasing availability and speed of electronic computing devices. A bad effect, as we have seen, is that they encourage us to over-correlate data. We can throw the data into a machine, crank out constants and then publish the values in the literature with the implication that they must be correct since they were obtained mechanically from a computer.

In the past when we wanted to design something we were limited to a few simple models which we could manipulate mathematically. Suddenly we are freed from this restriction. We can utilize any model no matter how complex either for correlation or for design. We need make no idealizations and can arrive at a final result which is limited in validity only by the experimental foundation of our model. Computers should therefore pressure us to obtain better data and differential data rather than integral data.

I see that my time is up. I would like to conclude by reading a quotation from Mark Twain's "Life on the Mississippi" which is somewhat pertinent to our discussion this morning:

> In the space of one hundred and seventy-six years the lower Mississippi has shortened itself two hundred and forty-two miles. That is an average of a trifle over one mile and a third per year. Therefore, any calm person, who is not blind or idiotic, can see that in the Old Oolitic Silurian Period, just a million years ago next November, the lower Mississippi River was upward of one million, three hundred thousand miles long, and stuck out over the Gulf of Mexico like a fishing rod. And by the same token any person can see that seven hundred and forty-two years from now the lower Mississippi will be only a mile and three-quarters long, and Cairo and New Orleans will have joined their streets together, and be plodding comfortably along under a single mayor and a mutual board of aldermen. There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.

SCREENING SOLUTIONS TO UPGRADE QUALITY:
A NEW APPROACH TO PROBLEM SOLVING UNDER CONDITIONS OF UNCERTAINTY

Norman R. F. Maier

Professor of Psychology
The University of Michigan

SCREENING SOLUTIONS TO UPGRADE QUALITY:

A NEW APPROACH TO PROBLEM SOLVING UNDER CONDITIONS OF UNCERTAINTY[1,2]

Norman R. F. Maier

PREFACE

This paper seems to need a preface, and I take the liberty of incorporating one. The objective of the screening principles is to raise the level of problem solving discussions when decisions are needed and further information is not available. It is understood that human beings may engage in faulty thinking but when one cannot prove that such thinking is faulty, how does one prevent it? As a matter of fact even when proof is possible, it fails to deter many because of the human tendency to persist in a point of view.

The four principles presented in this paper are reduced from an original number of nine, each of which had three or more subdivisions, making a total of 38 steps. This reduction was necessary because the initial process was too complex to use. It was found, however, that many solutions could be eliminated by several of the steps and this made possible the present simplified process.

It is hoped that the principles discussed will be viewed in the light of a new direction in the method of working with group problem-solving. It is recognized that there is a need for refining the principles and exploring ways in which they can be tested by research. Since some of the principles may appear irrational from one frame of reference it is necessary to regard the principles as useful in upgrading a group's product rather than in assuring a "correct" or even a "best" decision.

Decision making under conditions of uncertainty according to Luce and Raiffa(1) is the condition in which the outcomes of the various alternatives are unknown. This condition is satisfied by the type of problem under consideration in the present study. However, there is also a difference. Decision making implies a given number of alternatives, whereas in problem solving the alternatives must be created(3). Thus problem solving involves both choice behavior and the finding or creating of alternatives. Our study, therefore, involves both uncertainty as to outcomes and uncertainty as to the nature and number of possible alternatives. This type of problem is too complex to be handled by the current type of models being tested in connection with the decision making process. The screening process, though less accurate and precise, is a new kind of approach to this complex type of problem.

---

1 This paper was written in conjunction with a research project on problem solving supported by grant number M-2704, from the National Institute of Mental Health, Public Health Service.

2 The author is indebted to Professors C. H. Coombs and E. Lowell Kelly for their critical comments and suggestions.

The potential importance of this approach is not merely one of improving the quality of the product of group thinking but also may lead to the exploration and analysis of the nature of the inference process. Human beings make many inferences, they have hunches and claim to have insights. These processes follow logical principles different from those that are pertinent to proof or deductive thinking. As a result one cannot use standard logic in evaluating the inferences made by living organisms, who by nature go beyond what is known.

THE NEED FOR A METHOD TO SCREEN SOLUTIONS

A deficiency in facts in a problem situation does not discourage attempts to find a solution. As a matter of fact such a free condition often leads to an abundance of solutions so that a new problem is presented - that of selecting one from among the many possibilities. Whenever the facts in a situation are inadequate, it follows that a good many of the solutions put forth cannot be logically rejected. If an evaluation is to be made, additional procedures must be developed.

Under such conditions it would be desirable either to obtain more information or to experimentally test the merits of various solutions. However, many of life's problems cannot await the evaluation of solutions by these methods so that, for periods of time, there are stages when people can reasonably disagree regarding the merits and demerits of alternate solutions. All might agree that the solutions will differ in value when more is known, but until then, they cannot agree on the relative merits of various solutions.

In science, conflicting theories are reduced in number and essential refinements introduced as research knowledge increases. The function of crucial experiments is to narrow the field. But even in this instance there often is disagreement over the nature of crucial experiments. This means that what one person regards as proof may be insufficient evidence for another. Thus the amount of observational data needed to convince Darwin of man's evolution was not convincing to many others. When a man makes an inference he takes a "leap in the dark" and makes some additional suppositions. We might say he "makes up" some new facts so that for him his viewpoint or solution is very plausible. If no one possesses facts to the contrary, his "make believe" facts are acceptable to him, while others may feel cool or even antagonistic toward them.

This disagreement among men regarding the merits of solutions becomes even more marked when we examine some of the day to day problems. Actions have to be taken even when the choice made cannot be justified by the facts or circumstances. These actions or solutions to problems may be viewed either as guesses or as hunches. If we consider guesses to be purely a chance selection, then it follows that people should be of equal ability in choosing solutions that will turn out to be effective. However, the term hunch implies something more than chance - it suggests a personal logic. People differ in their hunches and some have hunches that turn out right more often than chance could permit. If people differ in their ability to solve problems when hunches rather than facts primarily determine the solutions developed, it is possible that this difference represents the contribution of creativity. Our immediate objective, however, is to determine whether solutions can be evaluated by procedures that do not make use of experimentation, logical proof, or probability analyses. In other words, when

facts and proof cannot be used to select the best solution from a group of possibilities, what sort of methods or skills can operate to insure better than chance solutions?

The problem is further complicated by the fact that people prefer solutions that they played a part in formulating and this may lead to popular choices rather than ones that eventually may be supported by additional facts. Selecting wise solutions under these conditions introduce numerous subjective factors so that questions of utility, statistical probability and rational behavior are only a part of the total problem solving process.

At present the hypotheses of man-made machines will be probability predictions rather than hunches. Perhaps this is their greatest virtue. In the meantime it is of interest to determine whether solutions to problems, particularly those involving people, can be screened so as to upgrade the quality of a group discussion outcome.

Illustrative Case[3]

In order to describe the screening process it seems best to use a specific problem so that the solutions recommended can be evaluated. In working with this material, the solutions were obtained by dividing groups of 10 to 25 persons into teams of 3 or 4 persons each. The teams were asked to serve as consulting firms and were given the assignment of finding solutions. Procedures for obtaining additional information, such as talking to the personnel in the case, were not acceptable as solutions because these were a means to the final solution. Each team was limited to three action steps. The request for having solutions stated in terms of actions to be taken was included because some groups state solutions in terms of objectives if not carefully instructed.

The case of the Sewing Machine Operators was then presented. The situational description was read by the leader, after which two persons in the group were appointed to read the dialogue. The final step was to supply each team with a copy of the above mentioned materials and a copy of the plant lay-out shown in Figure 1. A half hour was allowed for obtaining consensus on three action steps.

The Situation:

For the past six months Bill has been the supervisor of a sewing room of 25 girls in a garment factory. The girls are all union members. Until recently, his main problem was that of getting out enough production. Three months ago, however, all sewing machine operations were changed over

---

3 Taken from Maier, (2) Psychology in Industry, pages 39-42.

Production Standard = 80 units per team
Quality Standard = 8% rejects
Problem girls: Pat, Nora, Fran, Bea, Marg and Alma

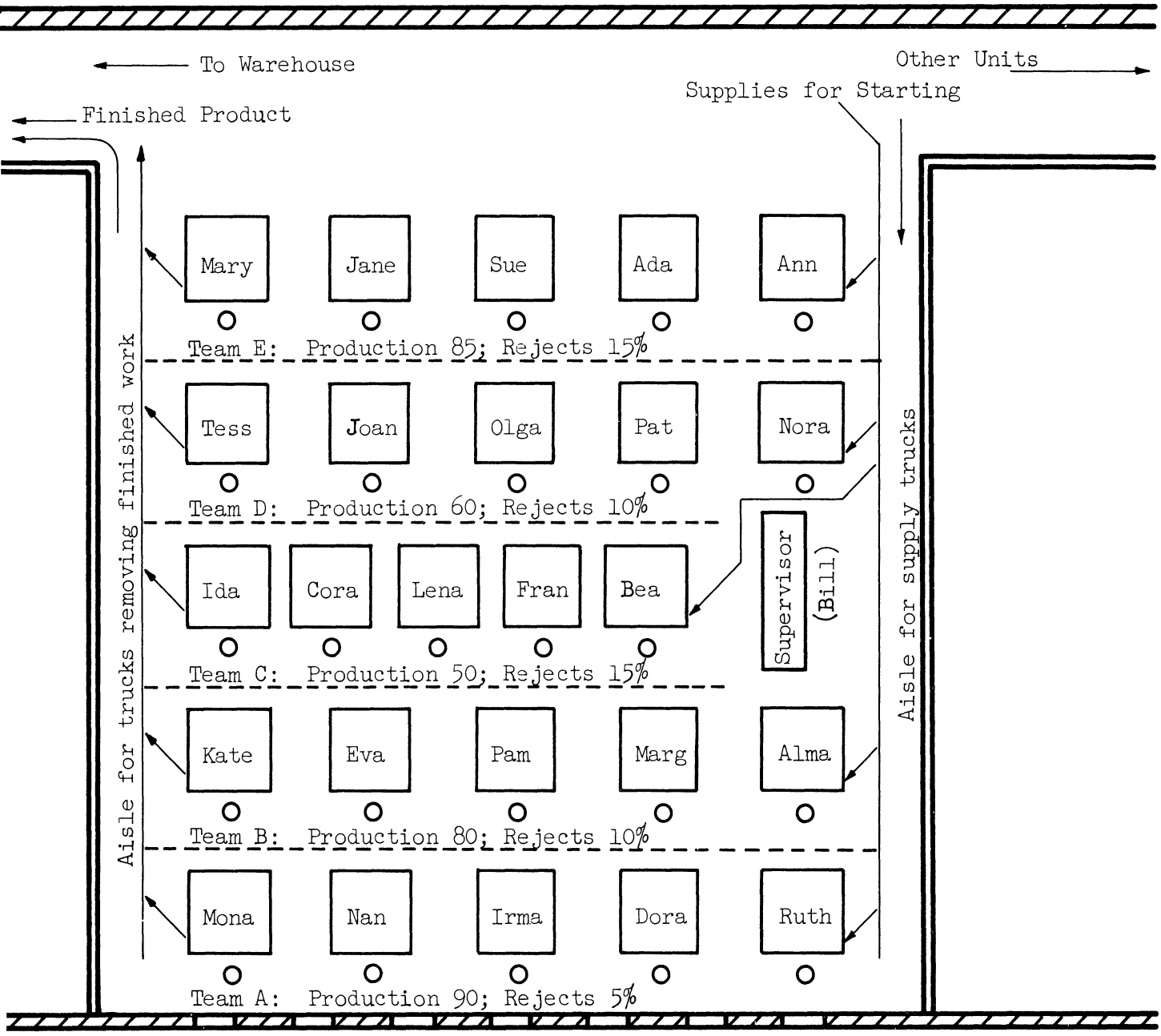


Figure 1. Sewing Room Layout.

from hourly to a team-piece-rate system so that production has become fairly satisfactory. Quality is now the big problem; not only are there too many rejects, but serious complaints are coming in from salesmen in the field. Since the girls are not paid for rejected items, it is difficult to understand why they are not more careful.

Bill has been called in by his department foreman, Mr. Johnson, to discuss the matter.

The Dialogue:

Mr. J: Bill, I want to talk to you again about the kind of work your unit is turning out. What's the matter down there anyway?

Bill: Darn it, I don't know. On the old hourly rate the girls weren't turning out anything, and now on this new group piece-rate a lot of the work they do isn't any good. When I make them do it over they say that I am picking on them.

Mr. J: Steve and Eddie aren't having the trouble with their girls that you do with yours.

Bill: Well, I'm not having trouble with all of mine. There is just this small group of five or six who are the real trouble makers. They all want to be finishers and anything but what they are. I've got them spotted next to my desk where I can keep an eye on them, and I tell them that I won't move them until they learn to cooperate. Even so, I'd like to see Steve or Eddie or anybody else get any work out of them.

Mr. J: You're not trying to tell me that just a few girls out of more than two dozen make your crew look that bad?

Bill: No, but they are the worst ones. I called all the girls in the "C" and "D" teams together last week and gave them a good talking to and now they're worse than ever. Production and quality are both down.

Mr. J: I'm beginning to think that you don't have any that are any good.

Bill: No, that's not right. I'll take the girls in my "A" team and put 'em against any we've got. As a matter of fact all of my finishers are a pretty decent bunch. The "B" team has a couple of good workers in it and there is nothing wrong with the "E" team.

Mr. J: Yeah, but their rejects are too high.

Bill: Well, that might be true, but those girls certainly produce. Maybe if I can get them to slow down a little the quality will go up. It's going to make them sore, though.

Mr. J: That's your problem. You're not afraid of them, are you?

Bill: No, but they didn't even like it the other day when I got after them for talking on the job. Come to think of it, all that gabbing may be the reason they don't pay any attention to quality.

Mr. J: Well, tell them that if they don't stop talking you'll break up their little club. You're the boss down there, aren't you?

Bill: Well, you've kind of got me there. I hired everybody in Team "E" in one batch with the understanding that they could work together and I hate to go back on my word.

Mr. J: Well, give them a good lecture and threaten to do it.

Bill: I know, but it's a headache and those girls stick together and you can't locate the trouble maker. Even when girls don't get along too well they gang up on you. For example, I gave these girls on Team "C" a safety lecture the other day after one of them got her hands caught, and all they did was gripe and pick on me about everything under the sun. I never saw such a bunch of sour hens before in my life.

Mr. J: What's eating them anyway? Certainly there must have been some one thing they picked on.

Bill: Oh, it was the same old yapping about all of them wanting to be finishers. After they've been on the job a few weeks they think they know everything.

Mr. J: Sounds like you've been giving some of those new girls a lot of half-baked ideas about the jobs around here. What's there to being a finisher anyway? The pay is the same.

Bill: I don't know. I think it's just a dumb idea they've got in their heads. You know how women are. I know the end job isn't any easier.

Mr. J: Was that all they griped about, or do they want us to give 'em the factory too?

Bill: No. Quite a few of them were sore because they said they couldn't make standard. Most of the girls think the bogey is too hard to hit anyway.

Mr. J: I haven't heard any other complaints about it. After all, 80 isn't so high. Why should your girls complain when none of the rest of them do?

Bill: All I know is that they do. Except for Team "A" and a few others who really turn the stuff out, they're just about the worst bunch of eight-balls I ever saw. I don't know why I have to have all of them.

Mr. J: Bill, we've been over all this before and I'm tired of listening to you feel sorry for yourself. Either get those girls on the ball, or we'll have to put somebody down there who knows how to run things. I don't want to be rough about it, but that's the way it is. I'll give you thirty days to get that mess straightened out, and I'll back you up on anything that seems reasonable. If you can show me some results by the end of that time you can stay; if you don't, we'll have to find something else for you. Is that clear?

Bill: I guess so. But after racking my brains like I have for the past six months, I don't know what you or I or anybody else can do with those girls. I've tried everything.

(End)

Examples of Solutions Recommended:

The following solutions are typical of those recommended when three solutions are requested. Solutions numbered 1, 2, 4, and 5 are suggested most frequently.

1. Turn lines 90° clockwise (also counter clockwise).
2. Move Bill's desk so as to give Team "C" more space.
3. Discharge Bill.
4. Train Bill in human relations skills.
5. Rotate girls within a team.
6. Rotate positions of teams.
7. Post production records of each team.
8. Obtain individual scores on rejects
9. Institute music programs.

10. Improve lighting.
11. Put in air conditioning.
12. Put five problem girls in one team.
13. Divide up problem girls and place one or two in each team.
14. Restrict talking in Team "E".
15. Break up Team "E".
16. Wall up back of room.
17. Train Johnson in human relations skills.

Screening Principles

It will readily be conceded that the 17 solutions listed are excessive in number and vary in excellence. The purpose of the screening process is to reduce this number by eliminating the ones least likely to be successful. In this manner the chances of taking action steps that turn out to be successful should be improved. This improvement in quality should be accomplished without sacrificing the group's acceptance of the selections made.

Two of the screening principles developed are designed to reduce the number of possible solutions by rejecting some of them, and two are designed to assist in the process of making a positive selection. Solutions that are neither rejected nor selected may be regarded as tabled.

It is necessary to discuss the logic underlying these principles in some detail because it is apparent that the screening process will not conform to the usual logical principles for evaluating a solution. The illustrations taken from the case material are not used to prove the merits of the logic but rather to demonstrate the way each principle can serve its purpose.

Negative Principle I. Solutions transferred from other problems should be rejected. Generally speaking it is wise to try to generalize learning, yet this principle states that solutions learned in another situation should not be transferred to this one. The reason for discouraging this type of solution is that it fails to solve a particular problem, but instead solves a class of problems. Acceptable solutions should be tailored to fit the known facts in the case. This does not imply that a solution borrowed from past learning will not be effective. It is granted that such solutions might be highly effective, but they also may be effective in problems other than the one under consideration. Such solutions then become solutions to a type of problem rather than unique remedies for a special condition.[4]

Although it may be argued that a general solution has greater merit than a specialized one, this insistence on a specialized solution has the following merits:

---

4 It goes without saying that generalized solutions apply to this particular problem when it is shown that the facts upon which the generalization is based are present in this problem. In such cases the facts in the present problem are used to support the solution.

1. It forces attention on the facts of the case.
2. It rejects pet solutions of group members while eliciting a minimum of defensive behavior.

The mere statement of a problem arouses ideas about what to do to correct the condition. In a group situation this tendency can lead to a considerable waste of time, not only because each person is inclined to make a contribution in the form of a solution, but also because the solutions become the core of the discussion. Thus one solution suggests others and soon the details of the initial problem cease to be examined. In rejecting solutions that are not founded on the facts of the problem, one directs the thinking of the group toward the known facts, even though these be limited in number.

People also have an emotional involvement in their own ideas. To reject a solution because it is a poor one leads to defensive behavior and, because the facts are meager, a good deal of time can be dissipated in face saving discussions. An arbitrary ground rule can screen out solutions in an impersonal manner by classifying them as "out of bounds" even though they might be regarded as good solutions. Extensive use of this principle has revealed that a person in a discussion group will readily concede that his solution violates a screening principle without becoming defensive.

Recommendations rejected by Negative Principle I are Solutions 7, 8, and 9. Solution 11 also might be suspected to be a borrowed solution and for this reason the persons recommending it should be asked to give their reasons. If their defense is based upon the success with the solution in other offices, then it is of the transfer type and must be rejected; but if the reason given is based upon the arrangement of the windows, it is allowed to be retained. For purposes of the present discussion we will assume it passes Negative Principle I.

Solutions 7, 8, and 9 usually stem from attempts to apply some knowledge about boredom or motivation to solving the problem and hence are of the transfer variety. Invariably persons suggesting them agree that the idea came from other situations.

Negative Principle II. Solutions, supported by facts or interpretations of facts that are challenged by other members of the group, should be rejected. It is assumed that each of the solutions not already rejected rests upon facts derived from the problem situation under consideration. Each solution, therefore, rests upon a particular selection or interpretation of these facts so that various solutions may have quite different supports. According to this principle, it follows that if a particular set of facts is seen as fictitious by someone else in the group or if a conflicting interpretation can be derived from the facts in the same situation, it is assumed that the solution rests upon an unsound support. Although a complete foundation is too much to expect in the type of problem situation

under consideration, it is reasonable to expect some unchallenged interpretations to support the solution. This does not mean that a particular interpretation must be proven sound - rather it must merely be better than those with conflicting interpretations or facts.

The merits of this principle reside in the discussion time saved by eliminating the debate over issues that cancel each other. Interpretations that are opposite and equal in strength not only consume excessive discussion time but also may be regarded as equally weak. The assumption is that high quality solutions would more probably be associated with unchallenged interpretations. It follows therefore that time saved in debating close issues could be well spent in the exploration of other facts and the examination of different interpretations.

Examples of solutions likely to be rejected by Negative Principle II are Solutions 1, 5, 6, 12, 13, 14, 15 and 16. If Solution 1 is supported by an interpretation that the window position is desirable because Team "A" is the best team, then a conflicting interpretation might be raised with respect to Team "E", which is farthest from the window. If it is argued that all teams should share certain advantages in position, then it is clear that this solution also makes all teams share the disadvantages that are present.

However, if it is argued that turning the lines 90° gives Team "C" more space, this solution should be retained unless it is found that the change takes space from some other team.

Solution 5 usually rests on the interpretation that girls want to be finishers (the last position). The fact, however, is that the girls merely <u>say</u> they want to be finishers; perhaps they just want to get away from Bill. However, either interpretation may be used to support the rotation plan. If this is done, then the conflicting interpretation is that rotation within the team places all girls closer to Bill just as much as it places all girls farther from Bill.

Solutions 6, 12 and 13 are subject to similar conflicting interpretations and for this reason are rejected.

Solutions 14, 15 and 16 are based upon the point raised about the talking of the girls in Team "E". The interpretation that talking is undesirable is based upon the poor record of rejects found for Team "E". It could be argued just as reasonably that talking is good, and this interpretation would be supported by the fact that Team "E" has a good production record. It is this conflict in interpretations that eliminates these three solutions.

Solutions 1, 2, 3, 4, 10, 11 and 17 remain after this screening, although 1 and 11 were questioned regarding some of their assumptions.

<u>Positive Principle I</u>. Solutions founded either upon (a) any of the unchallenged facts or (b) unchallenged interpretations of facts, taken from the problem situation, should be selected for consideration and evaluation. It also follows that, other things being equal, solutions that have

more than one fact or interpretation to support them should be considered to be more appropriate than solutions resting entirely on one of these supports. Furthermore, the more different the supporting facts or interpretations, the greater and more valuable the solution. For example, suppose that research in clinical psychology suggests that a nondirective counseling interview will cause patients to assume responsibility and that research in social psychology indicates that responsibility increases as the leader encourages participation. These two sources of data support the conclusion that participation makes for responsibility. A conclusion supported by such widely different sources of data would be regarded as more valuable than one based upon two related research studies, such as two studies on interviewing emotionally disturbed people.

We will now turn to the solutions that have not been rejected and determine whether any of them are favored by Positive Principle I. Solutions 2, 3, 4 and 17 can be used to illustrate how this principle selects solutions. Solution 2 has a basic fact to support it: Team "C" has the least space and it is the worst group. Although it can be argued that this relationship might be pure coincidence, such a conclusion does not challenge the fact- it merely is a statement indicating the need for more facts. Furthermore, the interpretation that the observed relationship between the work done and the space occupied by Team "C" is a coincidence, is supported by facts obtained elsewhere and hence violate Negative Principle I. The facts in this situation place the poor production record of Team "C" and its shortage of space together, and it is unlikely that anyone would seriously contend that Team "C" would do worse if it had more space. [5]

Some problem solvers attempt to support Solution 2 with additional facts or interpretations. For example, it may be claimed that the crowding of Team "C" creates problem girls. This interpretation can be challenged by the fact that Teams "B" and "D" have as many problem girls as Team "C", but have the full amount of space. Thus Solution 2 rests primarily on one observed negative relationship between space and productivity.

Solutions 3 and 4 are similar in that they make Bill the problem and as a result they are supported by identical facts. The following observations support these solutions:

a. Steve and Eddie (other supervisors) do not have as much trouble as Bill.
b. The actions that Bill has taken haven't been successful.
c. Finishers are the most satisfied and they are farthest from Bill.
d. The best teams are farthest from Bill.

---

5 It is frequently difficult to convince statistically minded persons to overlook the possibility that errors are produced by a failure to give full recognition to chance. Therefore, we wish to remind the reader that we are dealing with situations lacking adequate facts and no claim is made that Solution 2 is correct. It is merely selected and will have to compete with other solutions.

These interpretations tend to remain unchallenged even though it may be argued that the evidence is not conclusive.

Solution 17 is supported by the fact that Johnson has not been helpful to Bill. This inference usually is not questioned. Sometimes it is argued that training in higher management should precede the training of supervisors. However, this conclusion is based upon facts obtained elsewhere (Negative Principle I) and hence cannot be used to challenge an inference based upon the situation under study. Thus Solution 17 usually rests upon one support - Johnson's apparent inability to help Bill.

Positive Principle II. When exceptions to a trend in results can be satisfactorily explained, solutions based upon the trend should be selected for further consideration. Ordinarily an exception to a rule or to a trend in data is regarded as a weakness and hence might result in the rejection of a set of observations. According to Positive Principle II each exception should be examined to determine whether or not it may be associated with some other exceptional condition.

For example, it might be claimed that boys in the fifth grade in school can run faster than boys in the second grade. Let us suppose that in comparing 10 fifth grade boys with 10 second grade boys there are two exceptions to this claim in that two fifth grade boys were beaten in a race by all second grade boys. These exceptions would weaken the above claim even if they did not destroy it. Suppose, however, an examination of the exceptions revealed that the two fifth grade boys who lost the race were cripples. This finding would tend to reinstate confidence in the claim regarding the superior running ability of the fifth grade boys.

When the facts are limited, the appearance of exceptions may cause trends in observational data to be summarily rejected, particularly when no provision is made to give them special attention. Positive Principle II requires that the exceptions themselves must be defended if they are allowed to stand. Any fact in the situation that can explain why the exception occurs, therefore, becomes relevant to this analysis.

Solutions 10 and 11 usually rest upon trends in data. It is found that there is a general tendency for teams nearest the window to do the best. If we take the production figures for Teams A, B, C, D and E we find them to be 90, 80, 50, 60 and 85, respectively. Exceptions to a general downward trend are the figures 50 and 85. Thus Team "C" does too little and Team "E" does too much. The fact that three teams fit the trend, whereas two are exceptions, makes the claim for a trend weak. However, let us examine these two exceptions.

Team "C" with the lowest production, is exceptional in another way - it has the least work space. This last fact can explain this team's exceptional performance. The second exception concerns the production of Team "E" since this is too high, considering its position. This team also is exceptional in a second way - it is the only team made up of friends. This fact could explain the team's superior production in terms of morale. Granted that nothing

is definitely proven, it will be apparent that one exception might eliminate another.

Turning now to reject scores of the various teams, we may examine them in a similar manner. We find the figures for Teams A, B, C, D, and E to be 5, 10, 15, 10 and 15, respectively. In this instance the rejects of Team D may be seen as too low or those of Teams "C" and "E" as too high. Team "D" seems not to have an exceptional condition so cannot be explained away. The high rejects of Team "C" can be explained as before, but the high rejects of Team "E" now must be explained by some unfavorable factor, such as "working too fast". It is clear that attempts to explain these exceptions away are unconvincing, but it must also be granted that a contradictory inference cannot be drawn.

It is possible to argue that the windows serve a valuable function until the middle of the room is reached, but that beyond this point the distance from them ceases to be a factor. The strength of this argument will vary somewhat from one discussion group to another.

It will be apparent that the case for the trend in results would be greatly strengthened with only minor changes in the data. For the time being let us assume that even though the case for the trend is weak, the suggestion is present. Let us, therefore, see if a distinction can be made between Solutions 10 and 11.

Teams "C" and "E" have the poorest quality. Since Team "C" is poor in all respects this is not surprising, but Team "E's" contradictory performances must be explained. Since Team "E" has no problem girls and the production is good, it is more necessary to explain the high rejects than the high production. If this step is agreed upon then the next step can be taken. If this step is not agreed upon, then it follows that the trend in data has been rejected because it is necessary to assume that Team "E" is a good team in order to entertain the trend.

Let us assume for the purpose of further analysis that Team "E's" poor quality (high rejects) must be explained. Will poor lighting or poor ventilation explain the poor quality the better? Generally it is agreed that poor lighting accounts for Team "E's" poor quality more effectively than does poor ventilation.

Final Evaluation

The recommended actions that remain for further consideration are Solutions 2, 3, 4, 10 and 17. Since Solutions 2, 10 and 17 have one support each, while Solutions 3 and 4 have more than one, the latter two actions would be ruled as the ones most likely to produce results. However, Solutions 2 and 10 are relatively inexpensive and might also be acceptable for that reason. Furthermore, Solutions 4 and 17 would be related in that they could be

made a part of a single program. Thus the final decision rests with ways to integrate or combine selected solutions in a practical manner.

In working with the screening principles we have repeatedly been able to cause teams of problem solvers, who were in initial disagreement, to change their selection and settle for Solutions 4 and 10. Solution 3 usually is not accepted because blame is also placed on Johnson, his boss. Since Solutions 3 and 4 are alternatives, preference settles on Solution 4. This degree of agreement cannot be achieved through a general discussion of results. Such discussions drift in various directions and persons tend to defend their initial decisions. Solution 10 rarely survives when the screening principles are not used for selection purposes.

DISCUSSION AND SUMMARY

A good deal of discussion time is lost in group problem solving because of difficulties created during the process of trying to solve an initial problem. These difficulties constitute a source of new problems, which interfere with the solving of the original problem because they are purely artifacts created by interpersonal relations. Such factors as injured feelings, faulty communication, personal involvement, insecurity, and face-saving can function so as to necessitate compromise solutions or the acceptance of solutions that arouse the least opposition. As a result the products of group thinking often lack objective quality.

This unproductive aspect of problem solving is particularly disturbing when facts are inadequate and freedom of thinking is not restricted by the need to conform to known realities. Often a person will unknowingly advocate a pet solution for a given circumstance because it will solve another problem with which he is faced.

Although good conference leadership can reduce these sources of interference, it is apparent that there is a need for other methods for improving solution quality. Obviously an acceptable evaluation process must be impersonal. The following screening principles were developed to eliminate poor quality solutions and to select those with objective quality. The acceptance of this screening process to persons participating in group problem solving is evidenced by the fact that there is high agreement on merits of the solution that survive the screening process.

The two negative principles that screen out solutions may be stated as follows:

I. Solutions transferred from other problems should be rejected.

II. Solutions, supported by facts or by interpretations of facts that are challenged by other members of a group, should be rejected.

The two positive principles that select solutions for consideration may be stated as follows:

I. Solutions founded either upon (a) any of the unchallenged facts or (b) unchallenged interpretations of facts, taken from the problem situation, should be selected for consideration and evaluation.

II. When exceptions to a trend in results can be satisfactorily explained, solutions based upon the trend should be selected for further consideration.

Solutions that are selected by the positive principles are evaluated by group discussion in terms of (a) cost and practical considerations; (b) the way in which they may be integrated; (c) a selection from among alternatives when two actions are incompatible; (d) an examination of the extent of support given by facts or interpretations; and (e) acceptability to group members.

Solutions that are neither screened out nor selected are tabled and should be considered only when the positive principles have failed to select solutions.

It is apparent that the evaluation process makes use of group discussion, but only after the screening process has forced a thorough consideration of the facts in the situation. Facts cannot be used exclusively with the type of problems considered because they are deficient.

REFERENCES

1. Luce, R. D. and Raiffa, H. Games and Decisions: Introduction and Critical Survey. New York: John Wiley and Sons, Inc., 1957.

2. Maier, N. R. F. Psychology in Industry. Boston: Houghton Mifflin Company, 1955.

3. Maier, N. R. F. The Appraisal Interview: Objectives, Methods and Skills. New York: John Wiley and Sons, Inc., 1958.

APPROACH TO A PROBLEM IN MICROBIOLOGICAL CONVERSION

Durey H. Peterson

The Upjohn Company
Kalamazoo, Michigan

APPROACH TO A PROBLEM IN MICROBIOLOGICAL CONVERSION

Durey H. Peterson

May I point out to you that it is impossible to predict whether one has planned a decisive experiment. Only after the experiment is completed can one determine whether it was decisive or indecisive. There are, of course, many experiments which are not particularly decisive, and perhaps more are indecisive than decisive as these speakers have pointed out. But isn't it true that there is some value to such experiments anyway. You remember what Kettering said -- "at least we know a thousand things that do not work." In spite of this, however, I do believe that good and careful planning can frequently lead one to the solution of a problem in a reasonable period of time.

Now I am showing a couple of designs (Figure 1) for achieving the objective and I intended to give examples of each design. These examples, of course, are all in retrospect of some of the problems which we were able to solve by careful planning. For example, as I go along, I will just review briefly the rough part of our solution and then go on to details. We were careful that we did no fermentations of steroids until we first had a method for detecting small amounts of these substances. I believe that one must have a method first and a good one, before one can design a good experiment. Perhaps that should be part of the experiment itself. But here, for example, we used a very simple filter paper strip technique to screen some of these steroids which we are going to talk to you about. Then, we had to investigate the abilities of various microorganisms to carry out this particular reaction which I will describe,so that we had to examine several different microorganisms (See Figure 1A). We soon found one, but we found one that did not do

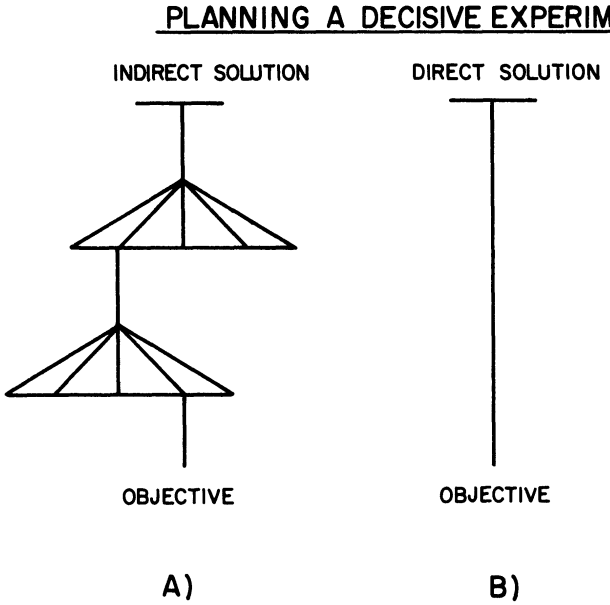


Figure 1

as good a job as we thought it should do, so we immediately screened this group of microorganisms (the Rhizopus sp.) and came out with a solution in which case we were able to get essentially quantitative yields of the product that we were trying to find. This is an example of an indirect solution to a problem as shown in Figure A. Now rarely do you find a direct type of a solution, but I do have an example of that, which I will give you later. (See Figure 1B.)

In order to discuss some examples of our work, it is necessary for me first to acquaint you with our basic problem, that is the structure of the steroids and specifically the structure of cortisone and hydrocortisone as well as a short history on the development of its use and its preparation. I do not intend to confuse you too much with formulas, but it is absolutely necessary that I give you some in order to acquaint you with the problem. In 1949, Kendall and Hench at the Mayo Foundation in Rochester, Minnesota, made the dramatic observations that the administration of cortisone, or hydrocortisone to patients with rheumatoid arthritis would alleviate their conditions. As a matter of fact, these workers had patients who were not able to get out of bed or even take one step; however, within one week after treatment the patients could walk upstairs. At this particular time the main problem became one of introducing the oxygen into the carbon 11 position (Figure 2). It is absolutely necessary that this oxygen be there for any biological activity. There were additional well known chemical methods capable of introducing hydroxyl groups at positions 17α- and 21- and thus produce an active molecule in the treatment of rheumatoid arthritis. In the early work it first required twelve steps to shift the oxygen from carbon 12 to carbon 11 in desoxycholic acid, which is a steroid acid found in ox bile. In addition to this, there were twenty-five steps required to complete the structure of this molecule and to produce this material. This, of course, meant that cortisone and hydrocortisone were too expensive for any wide-spread medical use. It was at this point that we were attracted to the possibility of making a product which would be available to these arthritics at a cheaper price. The main problem of many of the universities as well as many of the pharmaceutical houses was to introduce this oxygen at carbon 11, by a much cheaper method. I approached this problem with the idea of using microorganisms since I had used microorganisms with steroids 13 years before, working with the late Dr. F.C. Koch, of the University of Chicago. I would like first to show you a model of the steroid molecule. Except for these two ends here (the side chain and the 3-ketone) this is what you see on Figure 2. The

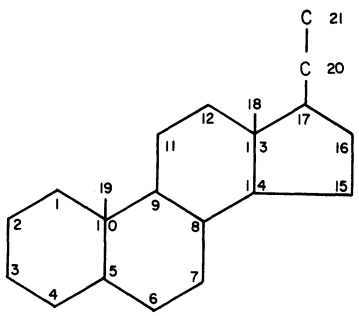
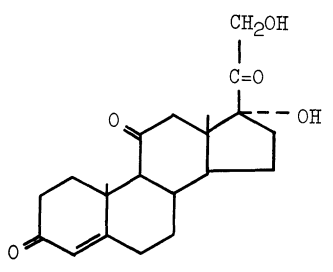


Figure 2.

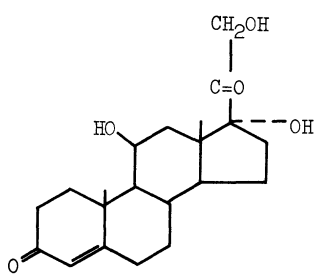
first three rings are six membered, and the fourth is a five-membered ring. Progesterone can be obtained in relatively cheap form, or from such natural products as diosgenin from yams, and stigmasterol from soybeans. These compounds can be easily converted by chemists to progesterone which can then be used for the introduction of the oxygen at carbon 11. Our microbiological approach to this problem proved outstandingly successful within a relatively short period of time. At the present time there are sold approximately $90,000,000 worth of these products, in which a good share of them are prepared by the microbiological introduction of oxygen at carbon 11.

These are the two compounds, cortisone and hydrocortisone, (Figure 3) which we aimed to produce, and there is the important oxygen in the form of what we call a ketone at position 11 or in the form of β-hydroxyl group at 11. These were the original compounds which Kendall and Hench used to do their work in the field of rheumatoid arthritis. So our problem became one of introducing the oxygen at carbon 11, as I have stated before, and in this particular molecule, that is on this carbon atom 11 in the figure. Before we did this work, however, it became clear to us that we had to find some kind of a rapid method for screening. It was absolutely essential to find such a procedure to determine whether addition of these cheaper steroids, of which progesterone is an example, could be easily converted to the 11-oxygenated compound. Progesterone has ketones at positions 3 and 20, a double bond in the 4, 5 position and then this acetone side chain at carbon 17.

We first used a microorganism which Dr. H. C. Murray picked up by exposing an agar plate to a window sill. By an agar plate I mean a petri dish which has certain sterile nutrients and on which microorganisms can grow. If one exposes such a sterile plate to the air, then incubate it, various microorganisms will grow. From that group, Murray picked out a mold which we were looking for, namely; a Rhizopus microorganism. We did this in order to study some of the molds which could be grown well and handled easily. The Rhizopus species are molds which grow on bread, as well as the mold Penicillium. When we did incubate this mold with progesterone, and which is written in that form on the figure (Figure 4A), we then were able to introduce this hydroxyl group in the 11-position. In other words, this is what happens (placing an α-OH group at carbon 11) . . . . and that is the way the molecule looks when you observe it as a model. I thought you might be interested in the following.



Cortisone



Hydrocortisone

Figure 3.

When I approached some of the experts in biochemistry, and chemistry, of this possibility, they said to me "look at this structure here -- you see this carbon 11 is all hemmed in by hydrogens and carbons. You don't have much of a chance of getting an oxygen in at that position." Since I had experience in the antibiotics field in which there were some very remarkable reactions produced by these enzymes, from microorganisms, I didn't pay any attention to these particular remarks and proceeded anyway. It was approximately eight months from the time we started the project to the time that we were able to solve the problem. There was a good reason why we obtained the solution to this problem in such a short time, and it is the following:

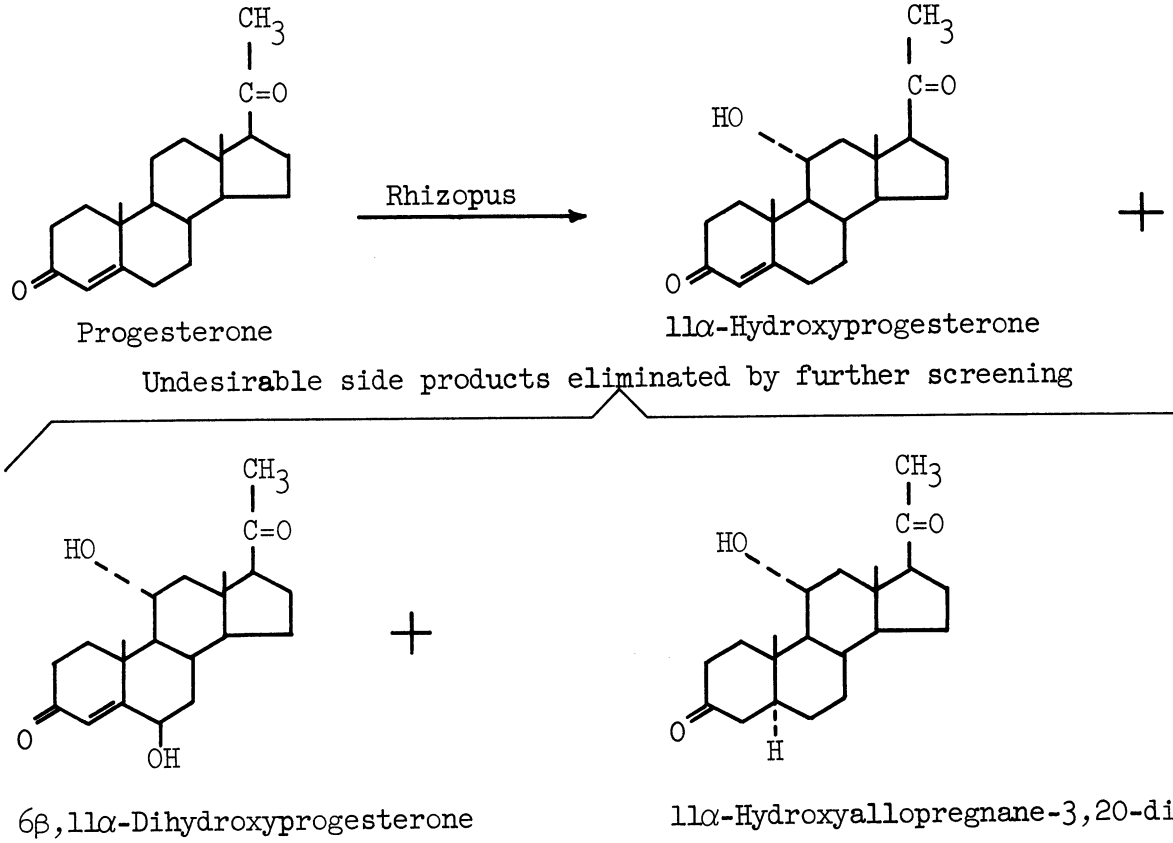


Figure 4

First, it was necessary that we had a good tool. We examined a paper strip chromatographic technique which I had worked on for the identification of various antibiotics and found that we could not apply those same solvents to the steroid field. In 1944, Martin and Singe, had invented this so-called paper strip chromatographic technique to differentiate amino acids. It was on this part of their work that they obtained, at least in part, the Nobel Prize. Just a simple piece of filter paper -- or blotting paper, if you want to describe it in a layman's language, was used very effectively on the amino acids. The early work of Martin and Singe was modified for use on steroids by Zaffaroni and Bush and this technique just happened to be available at the right time for us. Within a month after we started this project, we had this procedure effectively in use. I would like to describe it to you in the next figure.

Paper Strip Chromatography of Various Inks

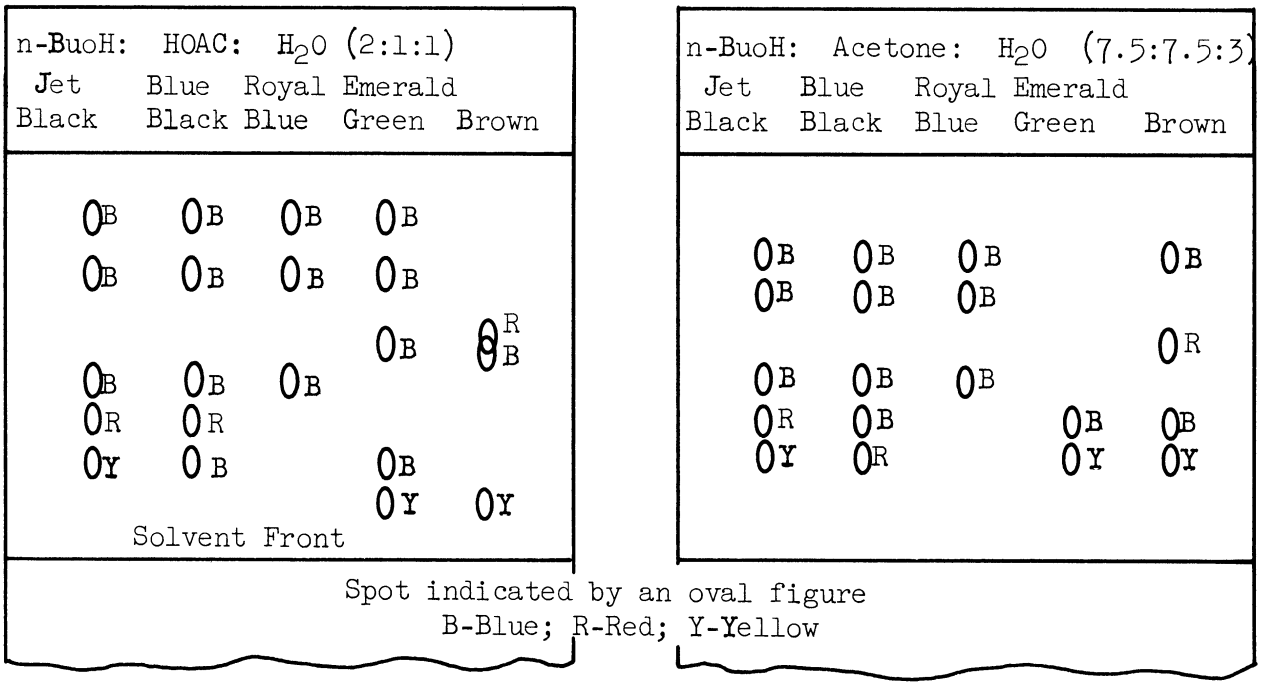


Figure 5

Briefly, I would like to tell you how this technique works -- let us examine the right hand side of the figure. One first spots along the upper line just a tiny drop of ink -- say a jet black, a blue-black, a royal blue, an emerald green, and a brown. If one then folds the upper part of the paper over into a solvent such as butanol-acetone-water, then capillary action proceeds very much like ink goes into a blotter. The solvent comes up this short fold of paper and then flows over the ink spots in a downward fashion. When the solvent flows over the ink spots, one can see that the inks are differentiated into various dyes or components. The jet black is a mixture of six different colors, the blue-blacks a mixture of several different dyes, etc. The left hand paper shows the same ink, but with a different solvent. In the same manner we can take colorless substances like steroids and do just exactly the same thing. We identify steroids after we have run these papers a certain period of time by different techniques. If they have the right structure the components can be seen under U. V. light. The compounds or spots can also be observed by spraying the sheets with various color agents. For example, once you spray them with a color reagent, they react, and form a yellow or blue derivative depending on the spraying agent used. Now each one of those spots are characteristic of a certain compound, which has moved down the sheet to a certain point and the position to which they move depends upon their chemical and physical characteristics.

Figure 6 shows an actual chromatogram in the original work where we used two milligrams of progesterone to do our first fermentation. We used 3/10 mg. of the product of fermentation to carry out this test and we had 1.7 mg. left. In order to show you how nicely the test works, over here in the next to the last vertical column is the material that we started with, namely, progesterone, which we wanted changed by these microorganisms. In other words this is just a control of untreated progesterone. In the first column we took the same progesterone and incubated it with a particular microorganism, Rhizopus nigricans. Instead of having much progesterone left, we have a new spot near the origin. This led us to scaling up this particular reaction, and isolating the 11α-hydroxyprogesterone. The latter is the compound in which we have now introduced the strategic biologically required oxygen, and obtain finally in quantitative yields. This compound can then be easily modified by the chemists to add the other necessary hydroxyl groups and produce the cortical hormones such as cortisone, hydrocortisone, and now many other derivatives of this type. This is an example of using several experiments to solve a problem.

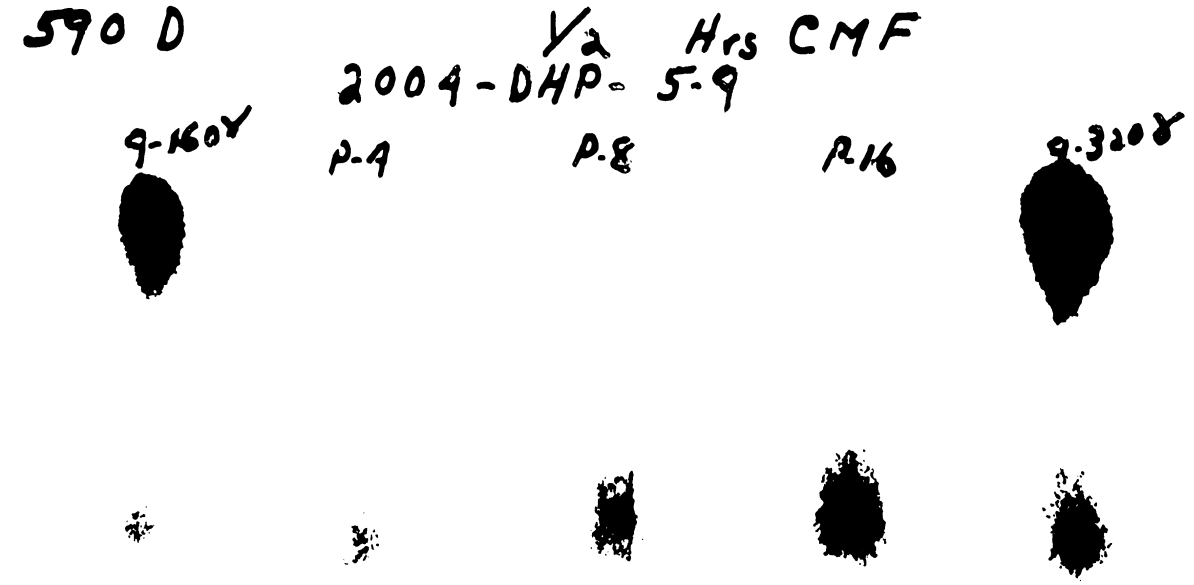


Figure 6

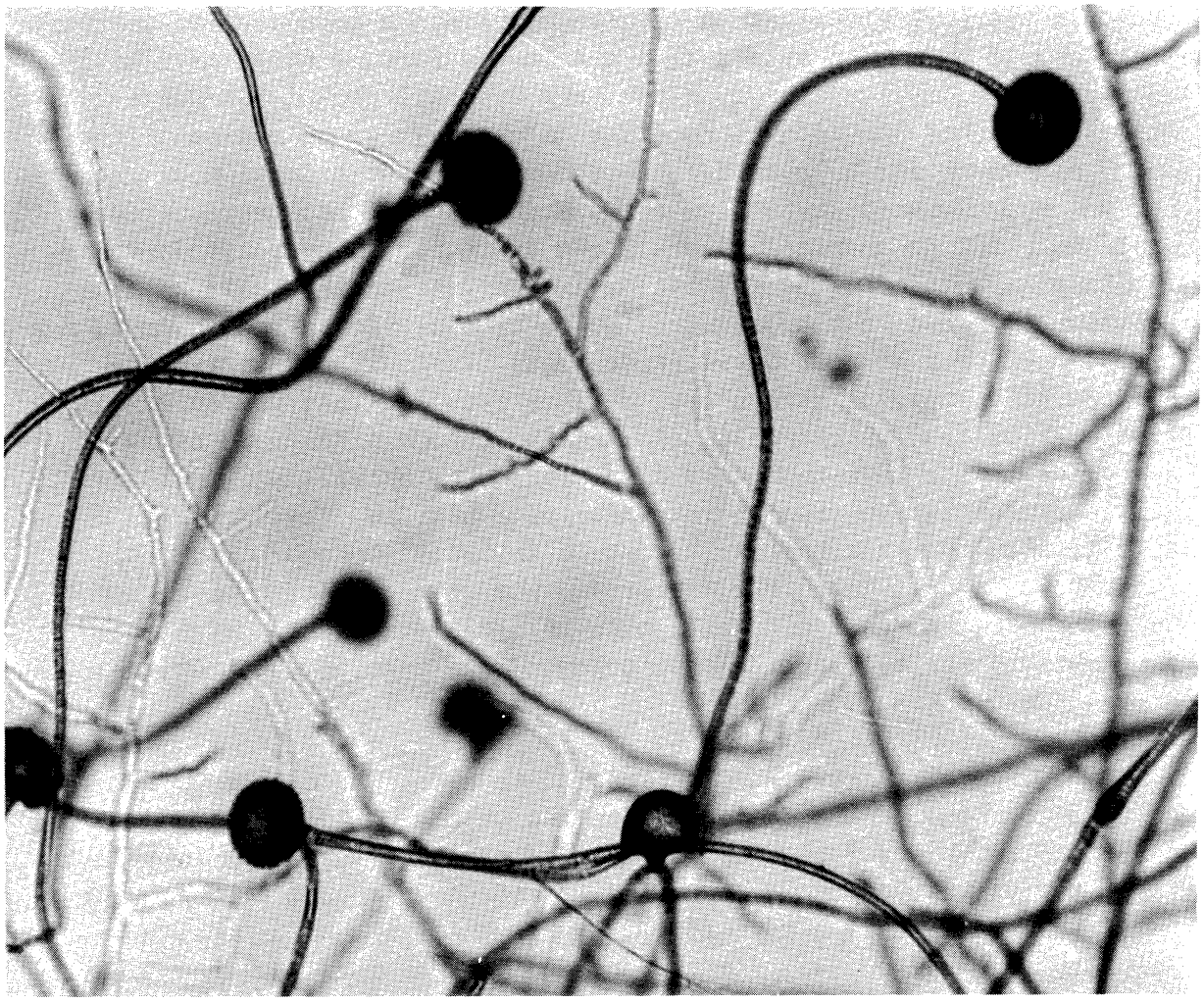


Figure 7

This figure shows you the biochemist in this reaction. This is a Rhizopus nigricans mold, a tiny creature which has the remarkable ability to carry out the 11-oxygenation of steroids. These, like all other microorganisms, secrete enzymes which are living organic catalysts that carry out intricate reactions while they themselves are not changed.

The following is an example of a direct solution and I don't think you run into this situation many times. Without regard to paying too much attention to structure, let us start out with compounds like Reichstein's Compound S, desoxycorticosterone (DOC), testosterone, or progesterone and incubate each compound with a particular microorganism. After isolation and purification of the product we found by analyses that we could, in each case, obtain a compound with an additional hydroxyl group. If any of you are chemists, one can tell by carbon and hydrogen analysis and by acetylation studies that the hydroxyl group has been introduced into a secondary or tertiary position. However, if an added oxygen in the form of a hydroxyl will not acetylate, it is probably a tertiary hydroxyl group. Then the question becomes "how do you know that a hydroxyl group has been introduced and where is it located?" We soon found that oxidation of the new monohydroxylated derivative of Compound S, DOC, and testosterone could be oxidized chemically to a known compound 14α-hydroxyandrostene-3, 17-dione. Therefore, the new hydroxyl must necessarily be in the 14α-position of these compounds. But in the case of the new monohydroxy derivative of progesterone it is exceedingly difficult to oxidize the acetone side chain at carbon 17. We then recalled that we had a mold (Penicillium sp.) which would oxidize the acetone side chain of progesterone to the 17-ketone. If this microorganism will also oxidize the side chain of the new monohydroxylated progesterone, then we would be able to show that the hydroxyl group is in that 14α- position. This was found to be true. Therefore, we proved the structure of this compound by using a microorganism. Well, this is a direct solution of a problem which I wanted to point out to you. This type of solution, of course, does not occur often.

There are several questions of interest which arise. In retrospect, some of the factors which led to the rapid solution of this problem then were: first the timely availability of an important tool, paper strip chromatography, came into the picture, because without it, we would have been a long, long time solving this problem. Secondly, I think one would ask, "why did you particularly choose a compound like progesterone?" There were two reasons, one was that our company happened to have a method worked out for the production of this

Structure Proof of 14α-Hydroxyprogesterone by use of Microorganism

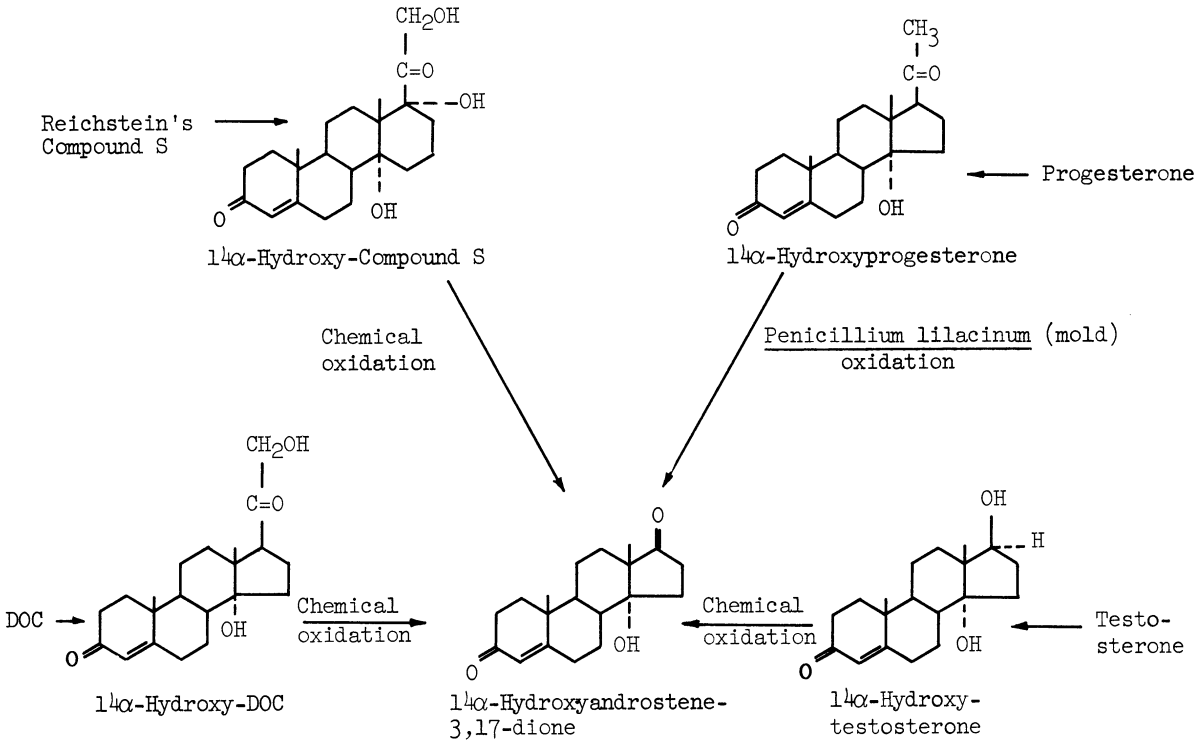


Figure 8

steroid because it could be prepared easily from an almost endless amount of soybean steroids, and secondly, I think perhaps just by a stroke of luck and the Grace of God, this basic starting material has turned out to be the best starting material of all these compounds. I thought you might be interested in hearing the story of a problem which was worked out successfully. I think that some of these experiments were very well-designed. However, one thing I failed to point out and that is that a lot of experiments we designed didn't work and I think you should know that too. We tried to plan good experiments, but even though they failed, they sometimes help lead one to the solution of a problem. I would like to acknowledge at this time H. C. Murray who worked with me originally on this problem, and also many other Upjohn workers too numerous to mention.

SOME TECHNIQUES FOR OPTIMIZING AN EXPERIMENTAL DESIGN

R. B. Morrison

Professor of Aeronautical Engineering
The University of Michigan

SOME TECHNIQUES FOR OPTIMIZING AN EXPERIMENTAL DESIGN

R. B. Morrison

You will notice that the title is one of optimizing the design, and not one of planning a decisive experiment which was the topic for today. I have deliberately changed the title because I think it is quite pertinent to look at experiments in terms of measurements that you perform. Thus, one might ask himself, "What proper measurements should I take to adequately define the phenomena that I am looking for?" Let me compare experiments to a paragraph in a book. If I pull out one word and say "Is that a good word?" without letting you read the whole paragraph, I have asked a nonsensical question. You should read the whole paragraph before you can look back and see whether that word is adequate or the best word for that paragraph. So you might look at all these measurements that you make to define an experiment the same as words that describe the sense of a paragraph, and what you would like to do in any optimization is to use the least number of words possible to describe the maximum information. Thus, much of what I have to say deals with measurements and their selection, but before going on into that, I would like to mention that experiments are usually performed for a reason.

One type of experiment is that of exploration. An exploratory researcher is an imaginative and creative person. It is hard to say to this man that this is the kind of instrumentation he should use, or the kind of experiment he should perform. He would probably rely more on his intuitive thought than he would rely on any prescribed technique. It would be impossible to fit him into a set of statistics and tell if he is a good researcher or not.

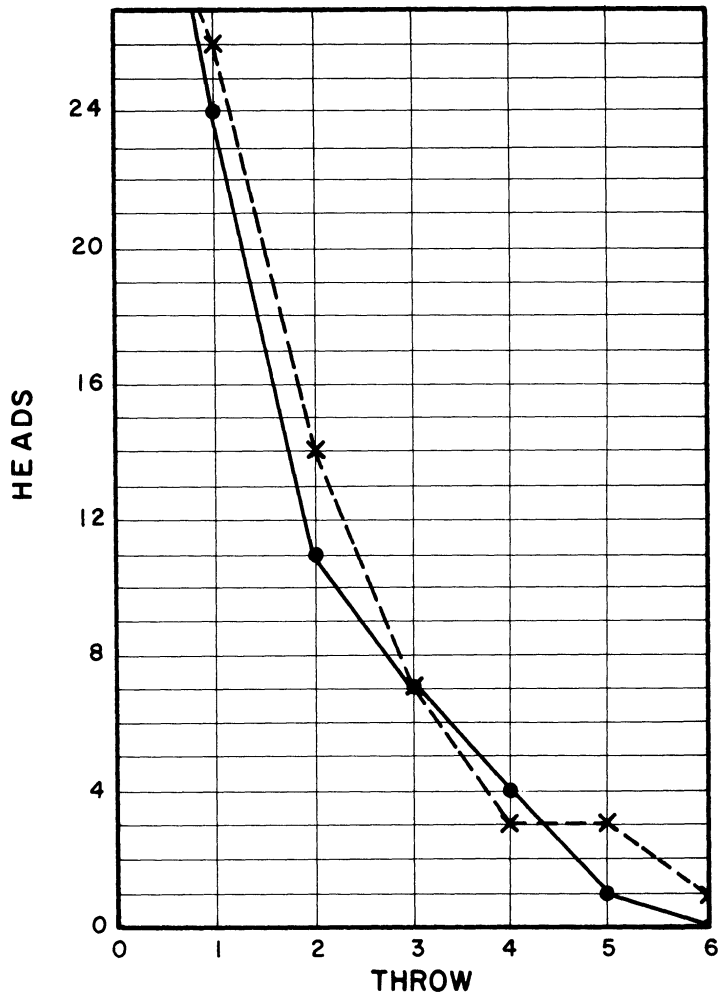
Another type of experiment is that of testing. This is in a routine type sense. In this case you are interested in satisfying some system specifications. Maybe it is a carburetor to see how well it works or what the limits are that it produces in the way of fuel-air ratio. Another type of data which is taken by experiment would be physical data, pressure-volume-temperature relations, etc., where essentially it is just pure measurement and the measurements in turn are presented in that form.

The experiments I would like to talk about deal with optimization type studies. These will go along with development-type tests. In this case, I would like to use development in a little looser way than it is generally thought of. In this sense we would use development as trying to improve or measure something to tell a story in advance.

Asking the Correct Question

It is very important that you ask the correct question. Let us take a box full of black and white balls. I could ask you several questions - "How many white balls are there in the box?", "What is the percentage of white balls in the box?", "Are there any white balls in the box?", or I could ask a complete question, "How many white balls and black balls are there in the box?" Each question requires a different procedure to find the answer. For example, if you wished to know the percentage of white balls in the box, I would immediately say "To what accuracy do you wish to know this?" This would determine how I would find the answer. If you want the best accuracy, count them. If you want to know within a few percent, and there are enough balls in the box, I could use some sort of photoelectric technique, on the basis that white balls reflect light better than black ones. There are all sorts of techniques; thus, I am sure that you can see that it is important to ask the proper question. There is a classic example in Morris' book on Operations Research, where he tells the story about the anti-aircraft guns on liberty ships in World War II. In this case, a high ranking officer asked, "Well, how many Nazi aircraft have we shot down with anti-aircraft guns on board liberty ships?" The answer was none. So he said, "O.K., take off the guns - they aren't doing the job." Somebody pointed out that the question should have been "How many ships without guns on board had been sunk in proportion to those that had anti-aircraft guns on board?" The answer was entirely different in this case. The percentage of ships sunk that had guns was much less than for those which did not have the guns. This obviously shows that the bombers were dropping their bombs further out on ships that had guns and their accuracy was therefore decreased. I had an experience at Cape Canaveral, Florida which I believe is another decent example of asking the proper question. This concerned gyroscopes which were sticking just prior to one of the missile launches. In this case there were about 130 gyros that were run through a series of tests and as they were tested more and more, more and more of them stuck. The situation became fairly critical so everybody got together and said, "Now look, we'll test these gyros and those that last longer than six or eight hours are probably good gyros, and therefore those are the only gyros we will use." Well, this isn't quite the proper solution is it? Failure in regard to time is not the only problem. In any event, we performed the above test and took the data of how many gyros had failed as a function of running time and obtained a curve of gyro failure as a function of time.

The sketch below represents a carefully controlled experiment that I did to help illustrate this test. This is like the story of the fellow who was taking the true-false examination. He was sitting flipping a coin and writing down his answers, when he got all through with his paper, he went back to the first page and started flipping the coin again and writing down the answers. The instructor came around and said, "What are you doing?", and the boy said, "I'm checking my answers".



Number of Heads vs. the Throw

Figure 1.

On the sketch it is seen that we started with 46 coins in a box. What we are really looking for in this case of constructive data gathering is two-headed coins (the one which is right all the time). We did this experiment in two runs, the solid line was the first run and the dash-line is where we checked our answers on the second run. We put the 46 coins in a box and shook up the box. We counted the heads (24) and rejected the tails. We shook the 24 coins and this time counted 11 heads. On the next shake there were 7 heads, etc.

Now in this case you can see that if you did have two-headed coins in the box, it would warp the distribution and you would come out high. Essentially that is the sort of thing that you would be looking for in the case of the sticking gyro. What happened in the case of the gyros was they were sent back to the company on the very gruesome note they were going to exclude them from the program, that their gyros were no good, and that they had better get busy and rebuild the faulty gyros into good gyros in a hurry or probably Eisenhower would be down there personally to take the company apart. This, of course, is the wrong attitude. The suggestion that we had was to take the same gyros that failed and put them back through a test and keep testing them until sufficient data was available for intelligent conclusions. The important thing was to determine whether or not the sticking in the gyros occurred at a random level, at a random position, or at random times. It turned out that indeed these gyros did stick at random levels, positions, and times. Thus, we put these results into a computer to examine what the chances were for the missile to fail from gyro malfunction and it came out something real small. Incidentally, the gyros they thought were good all got sent back, they all stuck too. But it was a question of asking really the correct question. You measure the stickiness you have in the gyro and then interpret this in terms of what you are trying to obtain. Indeed at this time it did not endanger the flight of any particular missile other than at some rather gross odds that were extremely in your favor. I believe that this covers the example, asking the proper question.

Limiting Instrumentation and Experiments

One of the things that we should do is to reduce the instrumentation and the number of experiments we must perform to obtain the desired results. The following example has been asked on some of the aptitude tests for high school seniors.

There are one-dozen balls; one of these balls is light or one is heavy. The question is, what is the least number of weighings on a simple balance that you can use to indicate which is light or heavy, and whether it is light or heavy. In this case, with a simple balance you can identify the ball, as to whether it is light or heavy, in three weighings. We found two ways that you can find these results. However, there may be more. The solution to one way begins by balancing four against four and holding four in reserve. If the balance tips one way or the other, the four balls in reserve are all standard balls. The standard balls can then be substituted for some of those on the balance, etc. If the balance does not tip, all balls on the balance

are standard and attention may be focused on the four in reserve. Off hand, I don't think you would expect that you could gather this much information out of such a small number of weighings and small amount of instrumentation. This technique is not very complicated but it does demonstrate a systematic way of optimizing systems.

Test Models

A test model will greatly reduce the complexity of the test that you are running. If you have built this model specifically for a test, you will have some definite data to start with. As an example, the following sketch of a shock tube illustrates the kind of things that you can do by assuming a test model and then checking by some simple instrumentation what goes on in the experiment.

For those of you who are not acquainted with a shock tube, let's say that initially we have a diaphragm x-x, some high pressure gas in the reservoir and low pressure gas downstream. Burst the diaphragm and you get gas running down the tube. The gas that was at the x-x face will form an interface and the shock will run out the front as schematically illustrated on the pressure-distance plot above. The rarefaction wave goes back up into the tube. Looking at the pressure plot (Figure 3) you would say, "Well, this is a pretty complex system - let me use a great number of pressure transducers and I'll tell you what is going on". I particularly picked this case, because the solution to it comes out in terms of characteristics and does not have a specific mathematical solution in the sense that I can give you the equation of that curve. I can draw that curve for you in terms of characteristics as accurately as you wish. All I have to do is to have a model which I understand. This is one which was worked out many years ago - the case of a centered rarefaction wave.

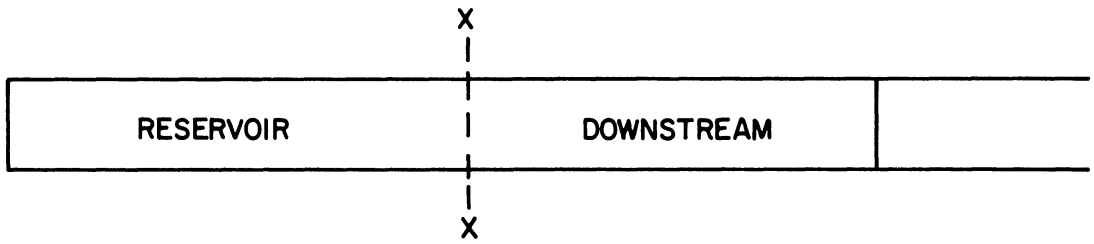


Figure 2

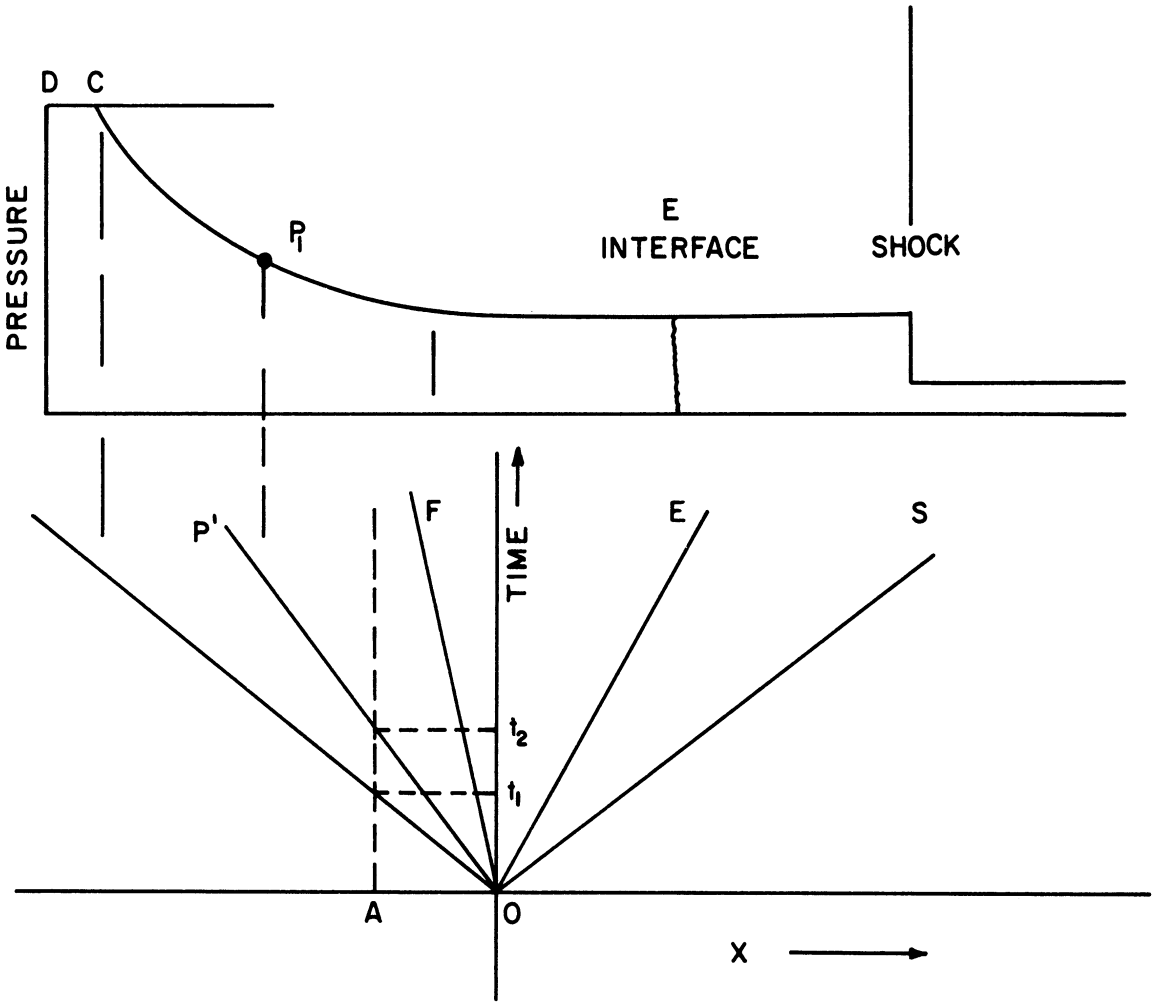


Figure 3.

In this case every property is constant along a straight line if you plot it as time versus distance. Take a pressure P_1 and follow its path with time. It is not the path of a particle, but just a locus of a point where the pressure is equal to some finite number and linear with time. The corner of the rarefaction wave, point C, is reflected at point D. The wavy band at E would be the interface location, which is represented on the time-distance plot as line OE. The shock location with time goes along line OS. The complete picture of what the rarefaction looks like can be deduced by taking one simple pressure measurement at A and following it with time and it traverses all of the pressures for you. If you have this model and you have assumed it correctly, then at the time t_1 you should measure the first rarefaction incidence, at the time t_2 you should measure the pressure P_1, and sometime later at a point whereof intersects the vertical line from A you measure the final pressure. This simple measurement at one point can give you, in view of having a model, a complete picture of what goes on in the reservoir section of the shock tube. It takes a little more than what was said here to realize that you can obtain the final pressure, but let it be categorically stated that you will be able to get this pressure as long as the corner of the rarefaction wave is moving to the left.

Another example is a model for a traffic intersection (Figure 4) which has been studied and which seems to offer quite a bit of promise.

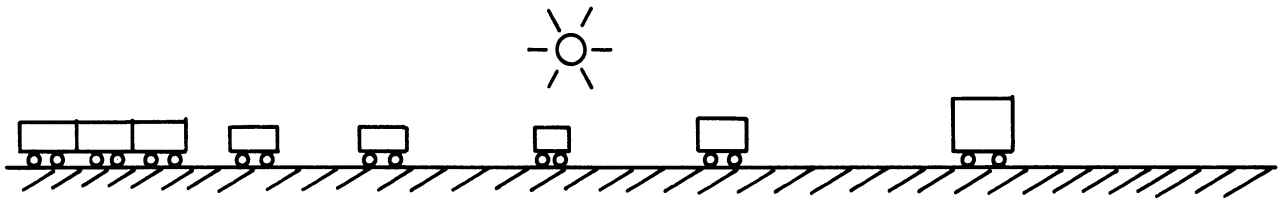


Figure 4.

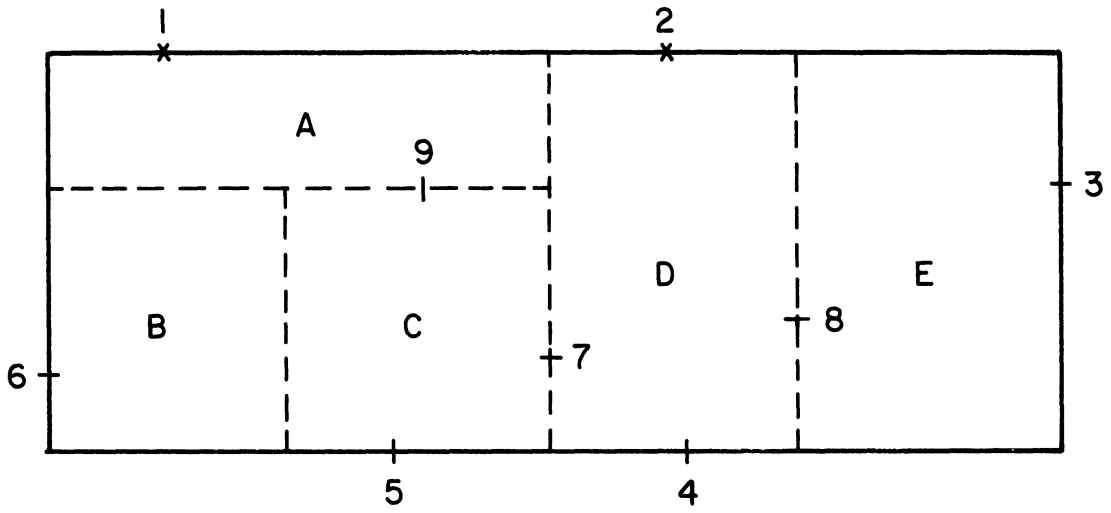
Several cars are stopped at a traffic light, the light turns greem and the cars go charging off. The question is in this case is that of the dynamics of the traffic. In identically the same way as in the case of the shock tube wherein pressure was a characteristic line, you can take a speed of 5 mph that a car is moving and it will likewise be a straight characteristic. More explicitly, instead of this being a pressure vs. time plot, let's look at the number of cars, or the velocity of the cars, versus time. Let me substitute for P, velocity, then OP, of would be straight lines and you will have characteristic, lines for the intersection problem. Not part of the talk, but very interesting, the speed at which the cars go through the light is constant and equal to a maximum flow speed (just like the maximum flow per unit area through an orifice) at something like 25 mph.

The conclusion reached from the two examples is that if you have assumed a model, put in a single measurement, run the experiment, and it agrees with what you know should theoretically happen, you have a decent certainty that you are on the right track. If the experiment does not work you know immediately you are on the wrong track. But in any case, you haven't spent a lot of time in instrumenting a test which is not what you want. You might say, "Well, how am I going to know what is going on?" Well, you don't unless you are successful, but I wonder if you could maybe not gather more information by thinking ahead, getting your model, putting in the minimum of instrumentation, and then see if your model is correct. Indeed, this is probably what is desired.

Optimizing Measurements

Optimizing measurements is in a sense the same as optimizing your experiment. The fewer the number of measurements necessary to describe what is happening in the experiment, the more optimum the design of that experiment.

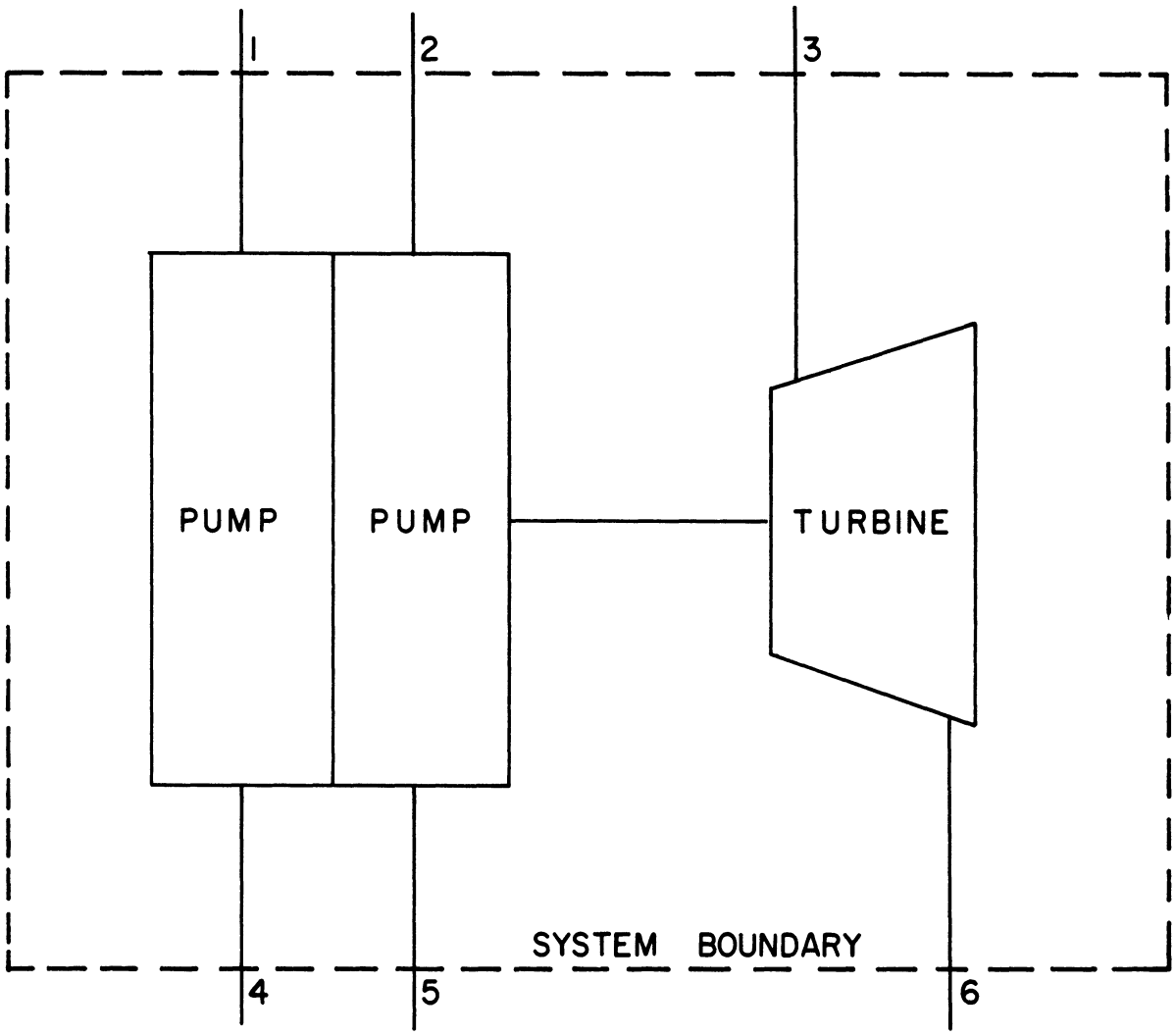
On the following figure (Figure 5), I would like to propose one thing by definition - that we classify measurements not according to priority, such as 1, 2, 3. As an alternate consider a system composed of several sub-systems, A, B, C, D, and E. In thermodynamics, you always define your system and put a boundary around that system. I would like to propose that in the case of any experiment, it is probably a good idea to put a boundary around it and treat it just like a thermodynamic system. The system boundary being the heavy line which goes around Figure 5. Any measurement made on the boundary, 1, 2, 3, 4, 5, 6, is a system measurement. Anything you make inside,



Types of Measurements
1. System
2. Subsystem
3. Environmental
4. Other: Safety etc.

Figure 5.

like 7, 8, 9 would be a sub-system measurement. Then there are measurements that permeate the whole region like ambient temperature of the system; that would be an environmental measurement. There are other measurements classified as safety measurements (perhaps you do not want to get too much carbon monoxide in the room). Thus, you have definitely defined the types of measurements. The type of measurements you should get at any cost would be system measurements. Taking too many sub-system measurements indicates your ignorance about the component inside. If you must take measurements 7 and 9, etc. around the little sub-system C, then maybe you had better study it by itself rather than make the problem more severe by adding the complexities of the other systems with all of their interactions. Applying this to a particular system, I would like to take one that is fairly simple and that was actually used. That is the turbo-pump system shown on the next page.



Turbo-Pump System

Figure 6.

In the case of a missile, there is a turbine that is fed by hot gas. It in turn feeds power over to run the pumps where you have liquid oxygen and kerosene going through the pumps to feed the motors. Now you put a boundary around this turbo-pump system. In this case we will make measurements where the pipe crosses the boundary at 1 and 2 and where the pipe crosses the boundary at 4 and 5 and the same for the turbines, etc. What is going across the boundary in this case is fluid. To specify the state of any fluid, in general (if it is a pure substance) you can do it with two properties. Perhaps pressure and temperature-any two that you wish as long as there are not two phases. You must know mass flows through the pipe also.

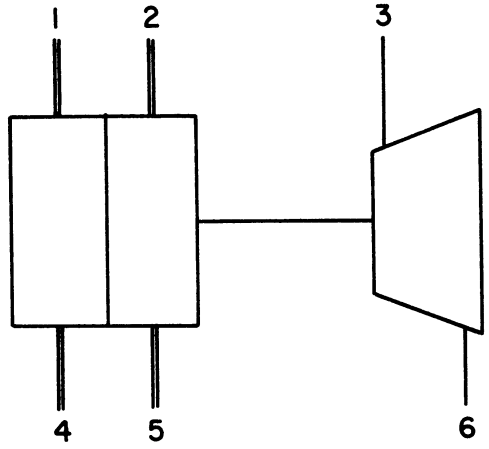


Figure 7

Measurements to Specify System Performance

P_1, T_1, $\mathring{W}_1$, P_2, T_2, $\mathring{W}_2$, P_3, T_3, $\mathring{W}_3$, P_4, T_4, $\mathring{W}_4$, P_5, T_5, $\mathring{W}_5$, P_6, T_6, $\mathring{W}_6$

noting $\mathring{W}_1 = \mathring{W}_4$, $\mathring{W}_2 = \mathring{W}_5$, $\mathring{W}_3 = \mathring{W}_6$

$T_1 \cong T_4$, $T_2 \cong T_5$

reduces measurements to:

P_1, T_1, $\mathring{W}_1$, P_2, T_2, $\mathring{W}_2$, P_3, T_3, $\mathring{W}_3$, P_4, P_5, P_6, T_6

or further noting that T_1, T_2, P_3, and P_6 not usually important to pump and turbine performance estimation

reduces measurements to:

P_1, $\mathring{W}_1$, P_2, $\mathring{W}_2$, T_3, $\mathring{W}_3$, P_4, P_5, T_6

In Figure 7, I have the list of measurements that you have to make if you wish to specify the pump completely. You would need to measure P_1 and say some other property T_1, any two properties would do, and a flow $\mathring{W}1$, then go on to 2, 3, and so forth. This would be the list of measurements that you would go through to completely specify the operation of that system. You will notice, also, that in doing it this way, there is the implication that if you know everything that is going in and out of the two pumps and everything that is going in and out of the turbine, that thermodynamics, with some analytical procedures, should connect the turbine system to the pump system, and vice-versa. By using this procedure, if one cannot analyze the system, then probably you have picked the improper system. In the above case, you can analyze the system. Let us see if we can simplify it some. Since you know that $\mathring{W}_1$ equals $\mathring{W}_4$, unless the pump blows a

casting, we can make that assumption. Note that the flow $\dot{W}_2$ is equal to the flow out $\dot{W}_5$ and the flow $\dot{W}_3$ equal at $\dot{W}_6$. For the liquid going through a pump in general, you could regard the temperature as being nearly constant, so we do not need more than one temperature measurement. Thus, you can simplify the method by reducing the measurements. Assume you are not interested in the temperatures of the liquids through a pump but the pressures. As far as the turbine is concerned, you should have an initial and final enthalpy as well as mass flow. Or if we let temperature reflect enthalpy, then we eliminate the pressures P_3, P_6 and have a final list of measurements which is about half of the original number. Nothing is degraded in this system. You have specified it within the assumptions we set forth, thus you can get a complete test on the unit and specify its performance. This is a good way of handling the problem.

A more typical way of specifying measurements, is the following example that has been used on missile programs. The propulsion people, air frame people, etc. are asked to specify the measurement's need. When results are totaled up, you come out with well over 2,000 measurements. Obviously, it is not feasible to make this many measurements, then all concerned personnel get together and by plebiscite decide which measurements should stay and which should go. This will total up to 800 measurements and telemetry has only a capability of about 150 measurements. The problem is immediately solved by taking the systems as defined and drawing a boundary around them. This results in about 150 measurements which specify most of the systems on the missile.

I feel it fairly important that, in experimental testing, time be taken to determine the exact measurements needed. In review, the steps are 1) define your system, 2) define the boundaries, 3) make measurements at the boundaries, 4) have a sufficient number of measurements to specify the state of the material going in and out (if it is electricity what are the currents, voltages, etc.), 5) examine your engineering analyses, to determine fluid requiring only pressure measurements, and 6) simplify your list. When you have done all of these things, you have a consistent list of measurements that will allow you to make desired performance analyses.

Figure 8 illustrates subdividing systems into subsystems. We have a system made up of subsystems A, B, C, D, and E. The numbers (.9, .8, .7 etc. in the first diagram) represent the reliabilities of the subsystems. If it is reliability you are checking for, i.e., a test you are running to determine the probabilities of failures, these divisionsare fairly poor. Looking at the two reliabilities .9 and .8, one can see that by multiplying them together you would have .72 and

could erase the boundary at A making that subsystem AB. Now let's subdivide subsystem D into 2 subsystems, D_1 and D_2, and we have approximately equal probabilities of failure within each of the subsystems. Now you have a system in which with a minimum of instrumentation you can locate trouble areas that arise.

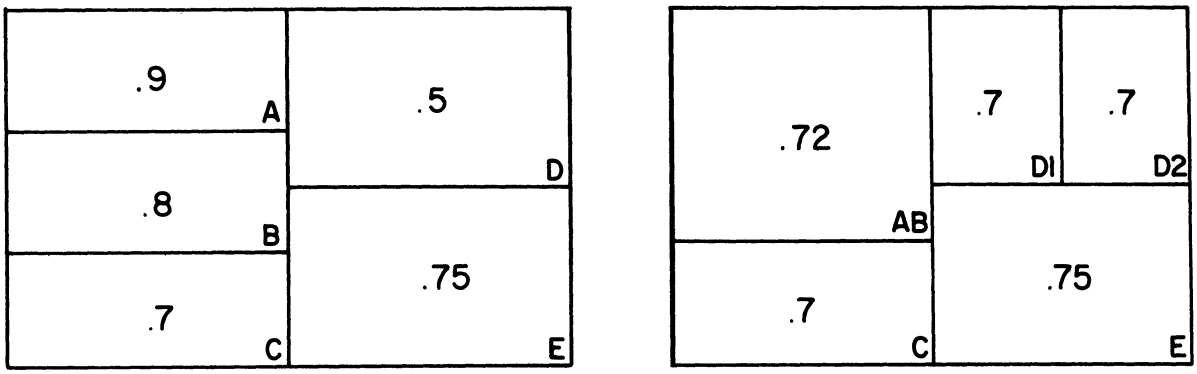


Figure 8. System Boundaries Established According to Reliability

Conclusion

To summarize this talk, I would repeat that you should look at your measurements as the source of your information. They should be looked at in a consistent and overall manner rather than individually. And like the words in a paragraph, the measurements are all you can believe, but probably you should doubt them and ask how accurate they are in giving you the desired information. Systems and system measurements specify your experiments, not single measurements except in very simple systems.

THE USE OF MATHEMATICAL MODELS IN PLANNING
AN EXPERIMENTAL PROGRAM

Wilson P. Tanner, Jr.

Associate Research Psychologist, Electronic Defense Group
The University of Michigan

THE USE OF MATHEMATICAL MODELS IN PLANNING AN EXPERIMENTAL PROGRAM

Wilson P. Tanner, Jr.

When Mr. Carroll called and asked me to talk at this meeting, the phrase "decisive experiment" looked like one about which you could very easily wind up and get very philosophic. I was quite puzzled as to what I should say and I worried considerably about this. I must have been in the right ball park, because its all been said already. At least I thought of the right things to say. So, under those conditions, I am going ahead and give the talk that the program looks like I'm going to give, without saying too much about decisive experiments as such.

I would like to repeat just a few words, however. The reason I would like to do this is because what I am going to do is sort of construct somewhat as Dr. Peterson did -- I'm going back a few years and construct an experimental program actually in psychoacoustics, and at this point I have the benefit of being able to look back and see what worked and what didn't work. As Dr. Peterson pointed out, certainly all of the things don't work. I guess one of the things you need if you want to be a good experimentalist is the ability to do an experiment and when you see it didn't work just toss it in the waste basket and not get upset over it. If you get upset over the ones that don't work, I don't think you will ever recover in time to do some that turn out properly.

Now I am not sure that there is such a thing as a decisive experiment. Certainly there isn't such a thing that can guarantee to be a decisive experiment in advance. You can get to work and you can state very carefully the hypotheses that you want to compare and you can design an experiment which certainly would lead to a decisive test provided that one of these hypotheses were good. There is no guarantee, however, that you've selected among the hypotheses which you have chosen, a good one. You may have selected a whole bunch of bad ones to compare. And your data is somewhere up in the middle and you can't really say very much about any of these hypotheses.

The other thing that I would like to point out is that there are at least two kinds of experiments. One is the type of experiment in which you are trying to compare different hypotheses and I think if there is such a thing as a decisive experiment, it falls in this class. The other type of experiment that you might be doing is one where you have decided to accept some model and you are doing the experiment merely to put parameters on this model so that you have a better description of the thing that you are studying.

Now when I speak of models, and I'm going to talk rather considerably about models, I think you should realize that the model that I am building first of all is merely a mathematical concept. Secondly, that even if I succeed in modifying it so that it fits the data, this doesn't mean that the model is actually any representation of what really is going on, beyond the fact that it is a mathematical equation, which will state the relation between the events that I have succeeded in observing. As a matter of fact, I intend to take a sort of probabalistic philosophy and these relations that I have observed are actually merely samples of some larger distribution and the relation that I actually think I have observed might just have resulted from a sampling error. You take enough data so that you hope this doesn't happen too often.

This morning Professor Henle talked about the theory of decisive experiments, and I detected in that a great deal of the theory of testing statistical hypotheses. The reason I mention that at this point is that I am going to talk about trying to conduct an experimental program to find out how a human being say, detects a signal which is embedded in noise. As soon as I embed the signal in noise, you see, and as soon as you try to construct an ideal observer for this task, this ideal observer actually is trying to test two statistical hypotheses. In other words, what we want to do, then, is analyze the situation of a person -- of anything for that matter -- trying to detect a signal embedded in noise in terms of this person who is trying to do the detecting. Consider him then as a statistical observer, or one who is testing a statistical hypothesis, and what we want then is a program where we can test hypotheses about this subject matter that we are dealing with, which is testing hypotheses.

The model I built is not only a model of what I think is a good description of the experimental situation that I am dealing with, but it should be, also, a good description of what I am trying to do.. that is, of the experiment.

The first figure is a block diagram which is, essentially, a block diagram of a device to test statistial hypotheses. I have essentially an input space -- the input space is multidimensional. The points in this space are derived from either a signal of some known variety, or no signal, and in both cases noise has been added to it. I put the first box here as a transducer. This is merely to call attention to the fact that we are frequently dealing with some sort of transformation of energy.

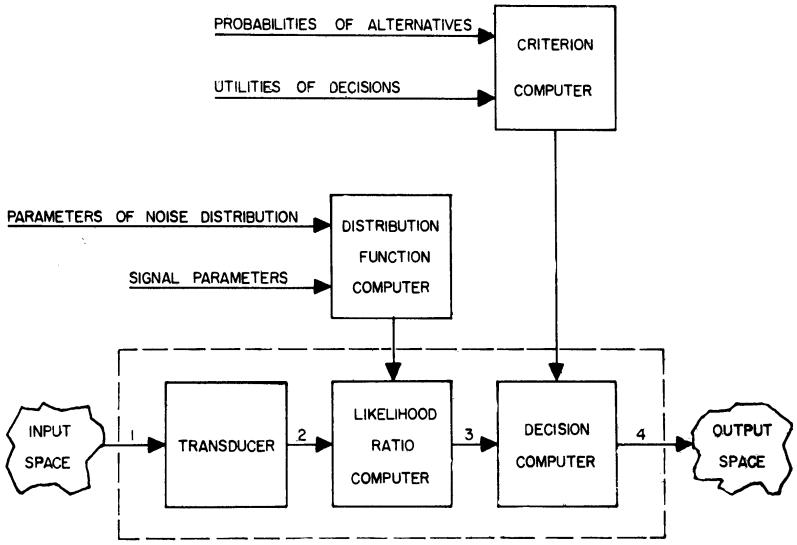


Figure 1

We do not want, however, to do any information processing at this point. The next box is a likelihood ratio computer. What you do is take the experimental data and you take the likelihood that if one of the hypotheses is true that this experimental data would arise to the likelihood that if the other hypothesis is true, that the experimental data would arise. This likelihood ratio computer then feeds a number into a decision computer and the decision computer matches this number into a criterion space. Once it is matched to a point in a criterion space, the decision is determined. It is at this point that the talker earlier today-that Professor Henle said that there was perhaps no rule that would give you an overall basis for accepting or rejecting a hypothesis in favor of another one. This is, in fact, true because you found that the rule for accepting or rejecting depends first of all on the choice of the function to be optimized. If you want to optimize something like expected value, you have one rule; if you want to optimize something like the preservation of information, you have another rule; and of course in the first case of optimizing expected value, you have a different rule for every particular situation that you might put into the thing. So there are many rules for accepting or rejecting as you please. Each rule being well specified as soon as you are willing to specify completely the various conditions which surround the situation, such that the payoff that you might get from accepting one of the hypotheses if it is true, and the punishment or cost that you receive for accepting this hypothesis when it is not true, and of course the payoff or the cost of accepting or rejecting the other hypothesis.

Now essentially the ideal statistical observer is sitting inside of this dotted line (Figure 1) and we see, first of all, that at this point there is certain information that is required for the likelihood ratio computer to do the type of operation that it needs to do. It needs to have the information regarding the statistical distributions surrounding the different hypotheses. By this I mean that we need to know whenever an input comes in, the likelihood that given one of the hypotheses that this input would arise to the likelihood given the other hypothesis that this input would arise. So we need to know the parameters of the noise distribution, the parameters of the signal distribution and these are then translated into statistical distribution functions. Up here you see we had the probabilities of the alternatives, and I have utilities of decisions. This is true of expected variant but is not true if it is information that you are trying to optimize. Now when I set up an experiment, say an auditory experiment, in which I am asking an observer to detect, say, the pulse of sine wave that is embedded in something like a white Gaussian noise, I can specify exactly the types of things that will happen if we have an ideal observer, and when I say the types of things that will happen, I mean on the average. If I have an infinitely large number of times that the observer is asked to state whether or not the signal is in the noise, and I have stated my a priori of probability, and so forth, I can state the probability that if the signal is there, it will be detected, and the probability that if it is not there, that the observer will accept it as being there, assuming that he is optimizing his behavior according to some decision rule.

You see what we've done is set up sort of an ideal observer which really gives us control of the experimental situation. What sort of a program have we begun to worry about here? Well, first of all we might be interested in such factors as to whether or not the observer we are dealing with can actually use the information regarding the signal parameters. In other words, is his performance better if he knows, say that this pulse is a pulse of a thousand cycle tone rather than if he knows that it is just a pulse of somewhere between a thousand cycle sine wave and a two thousand cycle sine wave. Does it help him in any way in his detection performance to know exactly what the frequency is? Will it help him in any way in his detection performance, say, to know that this is a signal which is a tenth of a second duration rather than a signal which may be of some duration which is statistically distributed. Does it help him to know the parameters of the noise? These are all questions that we might ask. We might also ask whether or not he can vary his behavior and some systematic way based upon information that we give him about the a priori probability of the signal existing. In other words will he behave differently if we present this signal only two times in ten rather than eight times in ten?

The problem is to set up experiments where we can begin, first of all, to test the feasibility of a model like this, only in reference to those things which it might have been. In other words, we don't just test the theory absolutely, but we test it with reference to other theories which might apply. We must then look for say some other theory to compare this to, and obviously the one you would like to compare it to is the one which is currently accepted, or has been accepted, and that would be the notion that the human being is a threshold device which, when you put an input in, something pops and the person hears it, or he fails to hear it. In other words, what that would say is that you would get certain types of behavior which would be consistent regardless of the input information that we give in these boxes up above (Figure 1). We want to establish a program around this type of question. In other words, we've got our model now and we want to compare our model to something else, but we also use our model to guard our own behavior. I might just make one passing remark that this set of experiments we did some years ago and we have recently moved into the stage where, we, at least, accept the theory. I don't know that anyone else does, but we do, and we are busy now trying to do the type of experiment in which we are essentially trying to fix parameters. In order to do this, it has been necessary to develop a rather complex type of experimental design which I can illustrate in the next figure.

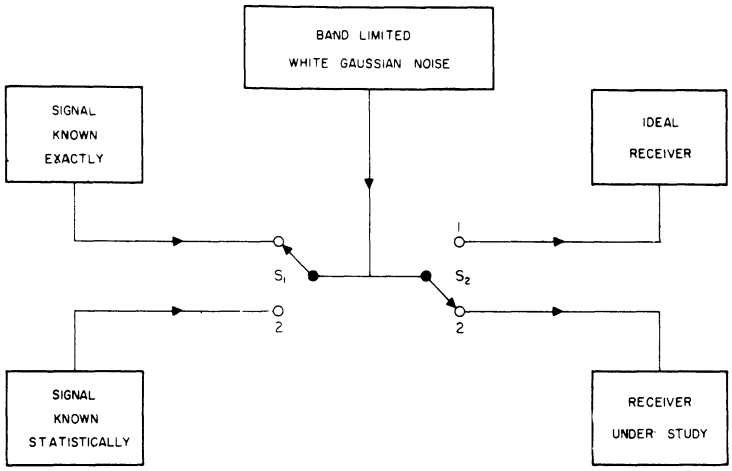


Figure 2

I will try to describe this in enough detail so you will get enough feel for it...it's complex and I might not be able to do this. If we have what we are calling here a signal specified exactly and it's embedded in an additive white Gaussian noise (now these terms aren't important at this point), we can state then in terms of the signal energy, and in terms of the noise power per unit bandwidth, the separation of the statistical hypotheses that an ideal receiver (and by ideal I mean one which is designed exactly to operate on that particular signal and

that particular noise) we can state, then, that the separation of the statistical hypthesis which that receiver can operate on. If we take this signal and, rather than have it known exactly (and by known exactly I mean in an interval 0 to T, lined up in exactly a particular way within this interval, as if it existed at all) particular amplitude, particular phase, and so forth. If I then take this signal and rather than make it known exactly, I give it statistical uncertainty as to its form, I find that what happens essentially is that these two statistical hypotheses would then be closer together. In other words, there is a smaller percentage of correct decisions here. This would be a signal known statistically again into an ideal receiver and in this case the ideal receiver is one which is designed particularly for this statistical ensemble. In order to use this to study human beings, we have had to develop the theorem that says essentially this. If we are using signal energy and we do an experiment with a signal known exactly through the first switch in position one, the second switch in position two, into the receiver that we are studying, and we get a measure of performance, then we do an experiment from this transmitter, switch one in position two and two now in position one, using the same energy, the same noise power, and we get an equal measure of performance this way, that this receiver because of its inability to use the information in the signal known exactly, has introduced into channel C_{12} the same degree of uncertainty that the transmitter has in the channel C_{21}. That's the theorem that we're using -- trying to find out how we can break down this ideal model that I showed you a few minutes ago such that we will make it match the receiver that we are trying to study. We have gotten, in this way, into some rather interesting studies on memory, which are just beginning to look at this point as if we can handle and come out with some statements as to the role of the memory in a human being trying to listen for signals. By memory I mean in this case how well can he essentially take this pulse of sound wave or whatever signal wave form we give him and store this in a way that he can correlate this with the input. The way that we do this is to design an experiment with provisions in the experiment itself which substitute for his memory. For example, if we add to this noise a steady sound wave component which is the same frequency as the signal he is trying to detect, the same frequency and phase, this signal then will just be an increment, but the fact that the steady component is there substitutes for his frequency and phase memory. And if we add the increment to a pulse, then the pulse being a large one is easily detectible,in most cases is one which carries starting time and duration information. I line this thing up in this interval so that he knows exactly when he should observe. I would like to make one statement, there that when we introduce enough different types of memory aids, we get observers who show an efficiency when compared to the mathematical ideal of nearly .5, which is quite a high figure.

What we've done with the model is that we've used it in a way which permits us then to study precisely what we're trying to do, the questions we're trying to ask of the experiments that we did. Using the model in this way helps us with our experimental design and we know which things to look for. If you are looking for a decisive experiment and essentially what you want is one with the two distributions pulled as far apart as possible, the farther these two distributions are pulled apart, the more likely you are to get a high likelihood ratio in favor of the one which is, in fact, the better one.

GOOD AND BAD RESEARCH

Peter Franken

Assistant Professor of Physics
The University of Michigan

GOOD AND BAD RESEARCH

Peter Franken

Let me begin by saying that there is no such thing as bad research. This is a sematic absurdity...you can have stimulating research, you can have ineffective research, you can have destructive research -- but there is no such thing, I think, as "bad research", and for the same reason there is probably no such thing as "good research". But we use the words frequently; I've even heard people say "good research comes out out of the University of Michigan". So it seems to be part of the accepted English abusage and I have no reluctance to use it in the title. Actually, the topic today is "Planning a Decisive Experiment". Most of the talks, at least from the abstracts and the ones that I've been able to hear, seem to deal with the methods -- the methodology of planning a decisive experiment -- which is certainly very appropriate to such a conference. What I would like to do is to talk a little bit about the people who plan a decisive experiment -- the people who do it -- in short, the researchers. They obviously play a vital role in the question of whether a decisive experiment is ever done. The title of this talk should probably have been "Good and Bad Researchers".

Now the way I would like to go about it -- I don't know very much about this subject -- is to refer primarily to my own experience. As advertised, I am a staff member of the University...I do Physics, and my primary activity lies in research. To this end I have a laboratory over in the Randall Physics Building, and on the average four or five graduate students working on their doctoral theses. Some leave, ...new ones come in. Usually they stay a number of years ---- the number tends to be classified.

When a graduate student enters the laboratory, he is raw. He may know a great deal ... he may have many skills. But he is not a researcher. And then when he leaves after awhile he is a researcher of some sorts; many of them are very good and are ready to go out into industrial laboratories and university laboratories and do effective work. All of these students have different attributes. Some are very gifted, very clever, some have only manual dexterity, others are just very creative without skills to back them up, and so on. It's hard to generalize...it's very hard to answer the question "what goes on here?" In comes something raw...out comes something less raw. I don't know the answer, but one thing I have observed, and I want to discuss today, is the difficulty that they all have, a difficulty that many of them never get over. For want of a better word I would call it the inhibition to do research.

Let me try to spell it out with an example: A student comes in and for the first three or four months in the laboratory he helps about (a sort of slave labor -- he enjoys this -- he becomes familiar with the ritual of the laboratory -- where things are -- he gets familiar with the program), and then after a few months he is given a problem. Go off and measure the microfine structure of doubly ionized lithium -- something nice and reasonable -- which he can tackle for a thesis problem. He is very happy about it -- he is given a bare room generally, and he is given funds to use, and he is very busy for a couple of months. But nothing happens. He comes in every day, he spends hours over catalogs, he orders things to be delivered in six or eight weeks. He makes many drawings for the machinists to be guided in fabricating his equipment, which again will be delivered in six or eight weeks. But nothing happens. You see, the thing the researcher (a good one) has to do and, this is the thing the students find so difficult to do, is to be able to come in a few times a week -- it doesn't happen every day -- come in in the morning and say "what can I do today -- what can I steal from next door -- borrow from upstairs -- beg from somewhere -- fabricate from string -- what can I do today as a sort of dry run to learn something?" Not six weeks from now, but today. These are terribly valuable things to do, because its out of these "dry-runs" that you very often discover some of the problems you thought were easy are not and that you are going to have some further things to worry about. Also, you often discover some of the things you are very worried about can be done very easily. For example, you may have been planning on having an ultra-high vacuum system because you are concerned about poisoning of surfaces or something, and you will sometimes discover by experiment that a water aspirator will do the trick. This has happened to me on two occasions. So its simple and direct to go and do something, but its terribly difficult. The student, particularly the beginning researcher, and many professional researchers I've known, are inhibited from doing this. It's somehow easier to come into the lab and say "what can I do now that is going to help me six or eight weeks from now?"

Another problem the graduate student often faces is a very curious one, but it sometimes can break them up. Finally after six or eight weeks, you see a lot of apparatus in the room; its all sitting in the corner, in the hall, and they've got to do something. They have to make a decision. Now the first decision is a very obvious one -- where to start building? Do you start in the center of the room? Do you start over in the corner by the sink? Or the other corner by the gas line. They find it is a very difficult decision (I've faced the same thing, too) ... it's a kind of writer's cramp -- they find it hard to realize that it doesn't make much difference. Start it anywhere -- start it on the opposite corner where there is nothing. It can always

be moved -- at least in many cases. In other words, get started. This inhibition ... it is very difficult to conquer.

Let me switch the frame of reference a little bit. One of the jobs I occasionally do, and it is occasionally rewarding, is to consult for several industrial laboratories. Now this consulting generally takes the following form. I'll spend the day there, visit in the morning and perhaps give a seminar on what we are doing here, and then spend the rest of the day walking around the laboratory, talking to the scientists, seeing what they are doing ... rubbernecking. And then inevitably around four o'clock I find myself closeted informally with some of the executive people in the company -- technical vice-president, research director, and so on. And what they want to know, quite reasonably, is what I think of the lab -- do I think there is good research going on -- or the possibility of good research, and if I don't think so, what would I recommend that they do about it? And so now when I go to visit companies, I expect this sort of thing -- I know around four o'clock it is going to happen. And as a result, I have tried over the years to develop rules of thumb to guide me. As I look through the lab and know I'm going to be asked questions at four o'clock, I keep my eyes on things. And almost inevitably these rules of thumb don't work -- except for one which I have found almost infallible. What I now look for, sort of out of the corner of the eye, is to see how neat the laboratory is -- how well organized it is. Let me sort of sketch the flavor of that. I look, say, at some of the capital equipment -- the mass spectrometers and things -- are they kept in reasonably good shape, or are they a little bit cruded up? Are the small items of capital equipment just stored on the workbenches, or are they filled away in nice steel cabinets. Are there hand tools scattered around the workbench, or are they all put away where they can be easily found? The reason I look for it is this: If a laboratory is neat and organized in the sense that I am describing, the chances, I have found, are almost overwhelming that there isn't a worthwhile thing going on. Now the reason for it, I think, is this. To be neat, to be organized is not just a simple matter of putting the debris into neat piles. It is something which takes effort ... it takes thought. It is a device that very often can keep you from doing something constructive, from doing research.

Let me try to illustrate this with a special example. Say a researcher is building -- here's his bench. He's building a piece of equipment as part of a research project. He needs the number four nutdriver, so he reaches up on the board and he gets it, and he does his operation. Now watch him ... does he reach up and put it back? Many of them do, and you ask why. This isn't just toilet training, gentleman. What he's doing, I think, is saying, "I want to stop what I'm

doing, break my pattern, put this thing back, because two weeks from now I might need the number four nutdriver and I want to be sure it's where I can get it." However, the good researcher, very often, puts it down on the table and goes on. He needs the pliers -- he reaches for them -- and so on. And now, there's no tragedy, for heavens sake. At the end of the day it takes five minutes -- he gathers them all up and puts them back in the drawer. But the consciously neat man, I think, is very often devising excuses not to do research. He's not excusing himself to other people -- its to himself, I think. He gives himself devices to provide the excuse for not devoting the full intensity of himself to the project. Research is a hard job. You don't do more than a few hours a day, but when you're doing it you don't want to have anything else on your mind, consciously, subconsciously. You're doing something. You don't have the time and you should not take the effort to put these things back and keep them well organized. I'm not saying, people, that you've got to be a slob to be a good researcher.

Let me take another example of the same problem. Most of the work we do, many of us, involves electronics. Not the sort of stuff you can buy, but the stuff that you have to build. The typical room in our lab has five, ten, or twenty circuits, each of them mounted on a panel, and five or six panels on a relay rack. One of them may be an amplifier, another an oscillator, and so on. Most of these are individually built. It's very interesting to watch a man build a circuit, particularly a man who's not a good researcher, because very often what he'll do is like so. He goes up to the stock room, and he gets a nice new steel panel with its black paint on it, which is a good way to begin. He goes to the shop, he lays out exactly where he's going to need the holes to mount the things carefully on it -- that's good, too. You have to take care with many electronic circuits. Now he begins to drill the holes. But he doesn't just drill them. He takes care. That is, he makes sure that there are not chips of steel lying on the drill bed press because, as it comes down, they would chatter and mar the finish. And the result of it is, after awhile, he'll end up with a beautiful circuit. It's gorgeous. The meters are right on center, the switches are lined up because he held them carefully before he tightened them, and it's a beautiful piece of equipment and it gives anyone pleasure to look at it. In fact, it didn't take him really any longer to do it that way than it would have taken him to do it more or less casually. But he has "ham-strung" himself in a very curious way. After a few months of building, he has a room full of these circuits and it looks gorgeous, and now he settles down and he is going to do his experiment. His Day of Reckoning has come. And inevitably as with all new equipment, it doesn't work. The panel with the amplifier on it, say; he discovers it doesn't amplify. Now what he should do, of course, is pick up an electric drill, go over

there, drill a few holes in front of it and mount that extra transformer that he needs. Just go over, drill a hole in it -- if there's a piece of metal in the way, cut it off, mount the new part on. But if he's got a neat beautiful laboratory that he's built up, he can no more go over there with an electric drill than he could take a pistol to his grandmother. He just can't do it. It's an amazing thing, but if you have a beautiful piece of equipment in which you've taken pride of construction, you can no more go and hack it up and modify it as a research instrument than the man in the moon. I've experienced this myself, and I'm sure many of you have. In fact, there is one thing that I intend to do more and more and urge my students to do also. When you build a circuit, hack it up a little bit ahead of time. This might seem to you a foolish thing. But I don't mean wanton hacking -- I mean this, you drill extra holes so you can stick tubes in in a hurry. Very often you just use chalk as labels, even because you may change them. And now you end up with a device that works just as well as the other one, but you have no committment to it. You can modify it. And this is terribly important. The man who has built neat, beautiful equipment is often stuck. What happens, you see, is that when he discovers he needs that new transformer put on the amplifier he turns his apparatus off. He takes the amplifier panel out, he walks up to the shop, he mounts it very carefully so he doesn't scratch the finish, and drills the hole and an hour later he's got it placed. But he has completely interrupted the train of his work, and besides which it's now 5 o'clock and he knocks off. It's bad in that sense. If he were free to hack into it, he would be that much further in engineering this inhibition to actually do research. Let me take one further aspect of this inhibition and then we'll have coffee.

I visited a large plant recently. It's one of many plants of a large international corporation. Two years ago it was a hillside and today there are 3500 people. Most of them are development and technical people. They are quite intimately connected with production, but they have a research laboratory. Ten or twelve scientists on the Ph.D. level -- all of this sort created overnight. Now the task of the research laboratory is to investigate any properties they want to of thin ferromagnetic films. These are right in the forefront of experimental work today, because they play a very valuable role in computer technology, memory banks, missles, and so on. These ten scientists were given all the funds they needed and quite a lot of space. This company does realize the advantages of completely unencumbered research. Everybody was very happy, but when I visited there about a year later NOTHING, absolutely NOTHING had been accomplished. It was a beautifully neat laboratory. The trouble in this case was quite striking, and it was quite clear at least, what it was connected with. These men came, and they had known each other before, I suspect, and they had all the

money they needed, so they started buying equipment. They bought a mass spectrometer because it's obvious that sometime in the future they would have to make an analysis of the chemical composition of the iron film. And they bought an infrared spectrometer for much the same reason. They might need it, of course, -- and a far infrared spectrometer, and a visible spectrometer. These are all $50,000 a crack ... also an ultraviolet spectrometer -- they got an electron microscope, ultracentrafuges, and so on ... because, after all, everyone needs an electron microscope, and an X-ray defraction camera. It's unbelievable. They spent $800,000. They had ten rooms full of beautiful equipment. They weren't doing a bit of research. It was full time work for ten Ph.D's just to optimize this equipment. One man, who is very good and has had a lot of experience, spent quite a while showing me how he had modified a $120,000 mass spectrometer so that it could give an analysis of the ferromagnetic material in 30 minutes instead of an hour. But he wasn't doing a single experiment. He had been ham-strung by this beautiful equipment. It's a curious effect. I don't know why he does it, or why the others do it, but its certainly again another facet of this inhibition to do research. I don't want you to misunderstand me on this point. I'm not saying to those of you who are research directors, or in that capacity, that when a man comes up and asks you for equipment you pat him on the head and say, "It's in your best interests that I don't give it to you." I'm suggesting this, though. When a person, one of your people, comes and asks you for a piece of equipment, whether its large or small, I think that if you are alert you can judge fairly quickly whether he wants it because he's in the middle of a project. If he wants a mass spectrometer because he <u>is</u> running an experiment on iron films and he needs this information ... then you get it for him as fast as you can. But if he's suffering this syndrome we were discussing, and he's buying it because he might need it, you might do very well to guide him in a somewhat different direction. But it's hard to do.

What I'd like to do now is sum up. Since I've been talking about this inhibition to do research, I would like to gather it all together in a neat package, tell you exactly what it is, and more importantly what you can do to help people who are suffering from it. I wish I could do this but I can't.

UNIVERSITY OF MICHIGAN
3 9015 03483 4773