GPO PRICE $ ____________

CFSTI PRICE(S) $ ____________

Hard copy (HC) 4.00

Microfiche (MF) 1.00

ff 653 July 65

Final Report
NASw 871

STATISTICAL DECISION PROBLEMS IN LARGE SCALE
BIOLOGICAL EXPERIMENTS

TO

Office of Space Sciences
National Aeronautics & Space Administration
Washington, D.C.

March 31, 1965

FACILITY FORM 602	
N66 29443 (ACCESSION NUMBER)	(THRU)
148 (PAGES)	2 (CODE)
CR-57677 (NASA CR OR TMX OR AD NUMBER)	04 (CATEGORY)

Bio-Sciences & Technology
Space & Life Systems Department
Hamilton Standard Division
United Aircraft Corporation
Windsor Locks, Connecticut

[redacted]

HAMILTON STANDARD
DIVISION OF UNITED AIRCRAFT

SLS 5305R

Final Report
NASw 871

STATISTICAL DECISION PROBLEMS IN LARGE SCALE BIOLOGICAL EXPERIMENTS

TO

Office of Space Sciences
National Aeronautics & Space Administration
Washington, D.C.

March 31, 1965

Submitted by:
D. R. Hitchcock, CoProgram Manager
Statistical Decision Program

G. B. Thomas, CoProgram Manager
Statistical Decision Program

Approved by:
R. B. Laning, Ass't to the Div. Vice President
Space & Life Systems Department

Bio-Sciences & Technology
Space & Life Systems Department
Hamilton Standard Division
United Aircraft Corporation
Windsor Locks, Connecticut

[redacted]

TABLE OF CONTENTS

1.0 INTRODUCTION - OBJECTIVES AND OVERVIEW OF ACCOMPLISHMENTS

The subject matter of the study conducted during the reporting period is the application of decision theoretic techniques to the design of biological experiments characterized by high cost, a relatively high degree of uncertainty about the correct functioning of the observational equipment, and a relative lack of antecedent information regarding the probable space-time distributions of the property or properties to be observed.

The objective of the now completed first phase of the study was to determine whether useful applications of statistical decision theory could be made to problems associated with the Martian biological exploration program.

Essential to the decision theoretic approach is the assumption that one is confronted with a set of alternative possible acts and one wishes to know which to choose, and that although one may not be able to predict with certainty what the outcome of each act may be, one can nevertheless specify the possible outcomes of each act. Associated with each act is an investment cost which must be paid if the act is chosen, and this cost is assumed to be known. Furthermore, it is assumed that one can assign to each outcome a utility, which is the value which would be realized by the decision maker if that outcome occurs. If the probability of occurrence of each outcome is known, then the <u>expected value</u> of each act may be expressed as the sum of the products of the utility of each outcome and its probability of occurrence given that the act is chosen.

Statistical decision theory concerns strategies for maximizing the expected value per unit investment in a variety of circumstances characterized by differences in the shapes of the utility functions defined over sets of outcomes and varying degrees of uncertainty regarding the probability of occurrence of each of the various outcomes.

In order to apply statistical decision theoretic techniques to real decision problems it is necessary to "structure" the latter by identifying the alternative decisions which could be made in terms of their costs and their associated possible outcomes, specifying the true or assumed probability distributions defined over each set of outcomes. It is also necessary to express in common terms the value of each of the outcomes of each of the acts. The decision problems which are traditionally most amenable to these approaches are those in which both the costs of the acts and the values of the outcomes can be expressed in terms of dollars, and where there is some way of validating, at least intuitively, the assumed probability distributions defined over the sets of possible outcomes.

The value of each of the possible results of an experiment undertaken to increase knowledge about Martian biology is not readily expressed in terms of financial rewards; although dollar costs can be associated with alternative experiments or experimental designs, the most important resources invested in them may not be a function of dollar cost but, rather, one of weight and size of the experimental apparatus itself, and the opportunity to send this experiment rather than another one to Mars on a given mission.

Prior to initiating this study it was proposed that these decision techniques might find useful application in making decisions about alternative designs for experiments to test a given hypothesis under the assumption that the importance of testing this hypothesis was already known. That is, we assumed that there was some agreement as to which were the most useful phenomena to study first, and supposed that the value of the information which could be provided by alternative experimental approaches might be expressible in terms of the probability that statistical inferences based on their results might be wrong in one of the two following ways:

The hypothesis under test might be rejected when it was in fact correct, or,

The hypothesis under test might be accepted when in fact it was false.

Early during the reporting period it was determined that the fundamental assumptions which the principal investigators had made were not valid. These assumptions are (1) that there was agreement as to the relative importance of the alternative experiments under consideration, and (2) that the objectives of each of these experiments could be characterized in terms of specific hypotheses to be tested or magnitudes to be estimated (and, consequently, procedures for interpreting the results of these experiments and relating the data gathered by them to inferences regarding Martian biota had been generated). Consequently, the study efforts were temporarily redirected toward what appeared to be a more useful and pressing decision problem, that of identifying and attempting to apply the criteria which ought to be employed in making a selection from an alternative set of available experiments from those which ought to be undertaken first. The opportunity to study this problem and to attempt to contribute toward its solution, was provided by Dr. George Hobby of the Space Sciences Division of Jet Propulsion Laboratories, who invited the principal investigators to participate in the JPL study of a Minimum Acceptable Biological Payload. A major portion of the study efforts conducted during the reporting period consisted of or derived from participation in this JPL study, the results of which have been published elsewhere. The primary output of this participation consisted of the development of an informal statistical decision model of the experimental selection problem which is reported in section 2 of this document.

Because the remaining studies completed or initiated during the reporting period were selected as a result of the conclusions formulated by the principal investigators during their participation in the Minimum Acceptable Biological Payload Study, the results of this participation will be briefly summarized here.*

1.1 JPL Study Participation

During their participation in the JPL study of a Minimum Acceptable Biological Payload, the principal investigators were concerned with explicating and critically reviewing procedures for evaluation alternative mixes of experiments which could be incorporated in a "minimum" payload,

*The opinions of the principal investigators and the conclusions they formulated as a result of the JPL participation should not be interpreted as necessarily reflecting the position of the JPL committee. This document has not been reviewed by Dr. Hobby. Furthermore, since much of the material was developed in part as a result of close collaboration with personnel of JPL, the material should also not be viewed as original contributions of the principal investigators.

and with comparing the relative advantages of "minimum" payload strategies with "maximum" payload (ABL) strategies. The informal decision theoretic model applied to these problems suggested a selection of experiments on the basis of the following desiderata.

1.1.1 The "cost" of each experiment as a function of the weight of the experimental equipment (including the weight of the required data transmission system and power supply) and the time and dollar cost of developing and testing the corresponding instrumentation.

1.1.2 The "utility" of particular experimental results as a function of the extent to which those results would increase knowledge about Martian biology, under the assumption that the experiments are designed so that the data provided by them can be unambiguously and reliably interpreted so as to support or reject relevant hypotheses regarding Martian biota.

1.1.3 The probability distributions defined over sets of results of experiments, as interpreted in terms of "a priori" probabilities which could be assigned to the occurrence of those results in the light of our present scanty knowledge of Mars.

These considerations cannot be explicitly and objectively employed in selecting experiments at this time because the available experiments have not been developed to a point which makes the evaluation of these factors possible. The objectives of the experiments cannot readily be characterized in terms of hypotheses to be tested or magnitudes to be estimated and the classes of inferences to which alternative sets of results might give rise cannot be readily specified. Similarly, the instruments to be employed have not been sufficiently developed to permit estimates of their "costs" in relevant terms of weight and power, or of their capabilities in terms of liability to specific types of error. Consequently, the principal investigators devoted some effort to analyzing certain specific classes of experiments in terms which would provide insight into problems of making such experiments reliable and determining the types of inferences to which their results might give rise and the kinds of costs associated with conducting them. These efforts included the following:

1.2 Analysis of Problems of Reliability

The reliability of an experiment is a function of many variables. One of the efforts completed during this reporting period consisted of reviewing many error sources and developing a theoretical presentation of the types of problems that might be encountered in controlling errors, with special emphasis on "errors of the third type," the seriousness of which primarily derives from the fact that experiments are conducted automatically in a remote location.

1.3 Simulations of Growth and Metabolism Experiments

Two simulations of growth and metabolism-type experiments were conducted in order to demonstrate the types of design problems resulting from definable kinds of "error," including machine error, measurement error, biological noise, communication error, bias, inoculum population variability, and growth rate variability.

1.4 Bandwidth Allocation Problems

If the weight of an experiment is limited and data transmission must be powered by internal power supplies (i.e., batteries), then the number of bits needed to transmit observations becomes an important cost factor. Since the value of such results in supporting relevant hypotheses is, in part, a function of how accurately they are known and the availability of information regarding the conditions under which these results were obtained, it is important to determine how best to use a limited bandwidth capability. Consequently, some attention was devoted to identifying the factors which must be taken into consideration in determining how many bits is enough and how many is too few.

1.5 Sterilization Policies and their Implementation

Certain policies regarding the likelihood of contamination of Mars resulting from the landing of inadequately sterilized spacecraft have been formulated. Sterilizing a spacecraft tends to degrade the performance of its components. Testing so as to insure that a given level of sterilization has been achieved by the sterilization procedures employed is expensive and time consuming. Thus achieving a given level of sterilization entails costs in the form of performance degradation, money and time, and alternative implementation procedures may vary with respect not only to these costs, but also in respect to their efficacy, which may be expressed in terms of the probability that a given policy will be satisfied. Some effort was devoted during this reporting period to explicating the factors which must be considered in evaluating sterilization procedures.

1.6 Analysis of Visual Experiments

Visual reconnaissance experiments designed to provide pictures on a microscopic or macroscopic scale are among the most appealing of the experiments proposed. Evaluating such experiments in terms which make them comparable to other exobiological experiments is not an easy task, primarily because it is difficult to express the functions of such experiments in terms of hypotheses which they can be said to test, or to compare alternative kinds of visual reconnaissance experiments in terms of the value with respect to the over-all success of the experiment contributed by alternative ways of implementing them. One of the efforts undertaken during the reporting period resulted in a specification of different ways in which visual reconnaissance experiments could be implemented and of certain of the factors which should be considered in comparing the probable scientific value of such alternative implementations.

1.7 Statistical Definitions of Life-Likeness

One of the most difficult problems associated with planning the Martian biological exploration program derives from the fact that life, if it exists on Mars, might not be recognizable as such because it exists in a form which does not occur on earth. Presently most life detection experiments are designed to detect Martian life only if they share certain specific properties with terrestrial life forms. One general property common to all terrestrial life is "negentropy" which in turn is expressed as orderliness. One of the activities undertaken during the reporting period was a joint effort with J. E. Lovelock* designed to attempt to develop some useful statistical tests which might be used to generate a measure of the amount of "order" present in a sample of Mars as a function of the likelihood of, e.g., the observed combination of chemical and physical properties in terms of which one or more experimental procedures might characterize that sample.

*University of Houston, Houston, Texas

2.0 DECISION THEORETIC APPROACH TO SELECTION OF EXTRATERRESTRIAL LIFE DETECTION STRATEGY

2.1 Background

The concepts presented in this document were developed while the authors were participating in a study conducted at Jet Propulsion Laboratories to define a minmum acceptable payload for the biological exploration of Mars. The objective of this study was to identify and recommend an initial scientific payload consisting of a set of instruments to be landed on Mars at an early date. A fairly wide variety of candidates for inclusion in this payload are "available" in the sense that they are currently under development or appear to be developable in time to include in the first payload. The problem was to identify that mix of experiments which seemed to represent the best bet for returning the greatest amount of useful information regarding the biota of Mars and which also satisfied the constraints implicit in the specification that the payload be a "minimum" one and that it be ready for launching during the '69 or '71 opportunities. Although the concept of a "minimum payload" was not a clearly defined one at the initiation of the study, it nevertheless was interpreted as being a payload which is small in volume and weight (and, as a consequence, of limited complexity both in respect to the number and variety of experiments which could be included and to the sophistication of any on-site data processing and control procedures that could be employed.) It was also supposed that a minimum payload should be the first in a sequence of scientific payloads. Although a very large number of considerations are relevant to the determination of the best mix of instruments to be included in the payload, one may usefully distinguish between those which concern the practicability of a given instrument as a piece of hardware, and those considerations which relate to the scientific value of the experiment in terms of the expected increase in knowledge which might result from including it in a payload. The practicability of an instrument is a function of its weight and size, the liklihood that an adequately tested flyable version will be ready in time, the cost of the equipment and its related development program, and the engineering reliability of the device. That an instrument be practicable in this sense is a necessary condition for its inclusion in an initial payload.

The scientific value of an experiment or an experimental strategy is less easy to define, but it is fundamentally a matter of how much a given observation or set of observations would increase knowledge of Martian biology and how likely it is that those observations will in fact result if the experiment is conducted.

The primary focus of the efforts of the principle investigators (and of the study group as a whole) was on questions of scientific value of alternative experiments and experimental strategies. Questions of engineering practicability received less attention at this time for two reasons:

1. Engineering data necessary for determining the practicability of the various candidate experiments was not available because none of the experiments had reached the requisite development state.

2. Strong claims regarding the scientific value of alternative experiments had been made by various participants in the program. Despite the fact that there is little uniformity of opinion regarding estimates of scientific value, decisions regarding the direction of the program appeared to be necessary, and it seemed as if these decisions would be based primarily on considerations of scientific value.

Although it was considered desirable that the JPL study group produce a set of recommendations which specified in as much detail as possible the components of the net electronic payload and the corresponding spacecraft/booster systems and mission profile, the objective was not a practical one given the lack of relevant hard data and manpower for such a design effort. More important, the basic decisions regarding the direction of the program for which this study was expected to provide some input did not concern the detail of a specific mission, but rather more far-reaching and general policy decisions regarding certain major cost and lead time items and, implicit in these decisions, the objectives of the early phase of the program, its cost, and duration. Thus, a basic question is that of the relationship between the scientific value of a payload and its size. The tasks assigned the JPL study group included the definition of a minimum acceptable biological payload and a comparison of this minimum payload with a highly integrated, automatic, non-minimum payload. Although completion of these tasks would obviously not provide a definitive answer to the question of the relationship between size and scientific value, it nevertheless was expected that a careful attempt to specify and evaluate a minimum payload would provide considerable insight into this relationship, and that, whether or not the recommendations of the study group were accepted, the process of critically reviewing these recommendations would at least clarify many aspects of this important question.

2.2 ROLE OF STATISTICAL DECISION THEORETIC CONCEPTS IN THE DESIGN OF A MINIMUM PAYLOAD.

The fundamental problems associated with selecting and defending a payload for the biological exploration of Mars arise from the fact that we know very little about the planet Mars and from the fact that contemporary biology, as a scientific discipline and a body of knowledge, is fundamentally earth-oriented and offers little guidance regarding the nature of the life forms which might be consistent with the Martian environment as that environment is inadequately understood. The limitations of available knowledge of Mars and the fact that, in view of the known differences between Mars and earth, most potentially useful extrapolations from earth-oriented biology are at best of questionable validity leaves the scientist who wishes to produce an accurate, or, at best, plausible evaluation of alternative experimental payloads with few firm grounds on which to base any conclusions. Conjecture, intuition, and opinion that is not readily supported by hard fact must of necessity play a major explicit role in arriving at any conclusions. Recognizing this fact, the members of the JPL study group felt that an attempt to apply certain concepts developed specifically for application to problems of decision making under uncertainty might prove beneficial, if not by providing formal decision mechanisms, at least to the extent of providing a set of concepts in terms of which this very complex problem might be usefully structured.

The primary advantage of this conceptual approach lies in the fact that the theory was developed as a technique for expressing and comparing alternative decisions whose consequences are not known with certainty. Essential to the decision theoretic approach is the assumption that one is confronted with a set of alternative possible acts and wishes to know which to choose, and that although one may not be able to predict with certainty what the outcome of each act may be, one can nevertheless specify the possible outcomes. Furthermore, it is assumed that one can assign to each outcome a utility, which is the value of the outcome to the decision maker. If the probability of each outcome is known, then the expected value of each act may be expressed as the sum of the products of the utility of each outcome and its probability of occurrence given that the act is chosen. Statistical decision theory concerns strategies for maximizing expected value in a variety of circumstances characterized by differences in the shapes of the utility functions defined over the sets of outcomes and varying degrees of uncertainty regarding the probability of occurrence of each of the various outcomes.

In an attempt to identify those frequently elusive factors which appear to contribute to the "scientific value" of an experiment, members of the study group analyzed the two primary experimental strategies under consideration and the types of arguments urged by individuals who advocate each. It was felt that a careful examination of these arguments might lead to the identification of those criteria which are employed in making judgements of relative scientific value.

The first of these strategies may be termed the direct approach to the question of whether or not life exists on Mars. It consists of focusing on specific biological features by conducting "life detection" experiments which would directly observe essential properties of life such as growth or reproduction, metabolism, morphology or would attempt to demonstrate the existence of life by observing biochemical constituents known to be essential to terrestrial life. The chief appeal of this approach lies in the fact that in terrestrial laboratories the detection of the presence of life forms is based on the observation of one or more of these properties and consequently the observation of similar properties on Mars would provide the only possible significant and conclusive evidence of life. Advocates of this approach recognize that Martian life may be so dissimiliar to terrestrial life that it either does not share these properties with terrestrial life forms, or they are present in a form which is not demonstratable by means analogous to familiar laboratory techniques. Such criticisms are countered by the observation that although Martian life may be different, we cannot either estimate the liklihood that it is different, not imagine what it would be like and hence could not design appropriate experiments to detect it; consequently the direct approach is best because no alternative approach that has any possibility of yielding conclusive evidence is available.

The other strategy which was considered is the indirect approach which consists of first surveying environmental properties of varying degrees of biological significance so as to provide a more complete though crude understanding of those physical features of Mars to which any life forms must have adapted, or those which might reflect the results of biological activity. The advocates of this approach are most concerned with the fact that Martian life forms may not exist at all or may exist in a form so dissimiliar to terrestrial life so as not to be directly demonstratable. Failure to demonstrate the presence of specific life forms will add little to our knowledge of what may be there. For this reason they feel that the search for specific life properties cannot be justified until we have better grounds for supposing that these properties exist and believe that an examination of environmental features is to be preferred because it can be guaranteed to provide some useable information even though that information may not be highly relevant to conclusions about Martian life.*

The study group considered a number of ways in which these two approaches might be implemented in terms of the types of observations which a first payload could be instrumented to make. This leads to a rough classification of observations in terms of whether or not they could

*This summary of the two approaches reflects the author's opinion that a fundamental difference between them lies in the interpretation of the objectives of the Martian biological exploration program: Implicit in the direct approach is the assumption that the objective is to detect life on Mars; the indirect approach assumes that the objective is to explore as far as possible all those features of the planet which do or would affect life on it -- including those which might preclude a biota. Although it might be plausibly argued that a program which satisfies one of these objectives necessarily satisfies the other, the question of the possible differences in the assumed objectives has not, to our knowledge, received explicit attention. It should be pointed out that we may be mistaken in our belief that the advocates of the two strategies differ in their interpretation of program objectives and in the priorities which are as a result assigned to various of the possible available experiments.

in principle provide conclusive evidence of Martian life, and, if not, in terms of the degree of biological significance of the inferences to which they might give rise. A careful analysis of the pro and con arguments associated with each of the strategies and with the corresponding classes of experiments resulted in the formulation of a set of scientific selection criteria which correspond to the considerations most frequently brought foreward in citing the advantages and disadvantages of different ways in which the search for extraterrestrial life may be implemented. This exercise resulted in a somewhat redundant (i.e. overlapping) set of six factors which are most often cited in arguments about the relative scientific merit of alternative experiments. Because they are often referenced in connection with the experiment selection problem, these factors may be provisionally referred to as scientific selection criteria. (That they in fact turn out frequently to be unuseable as such, and therefore ought not to be relied upon is one of the conclusions resulting from the particular model of the experimental selection problem developed). Their chief value lies in the fact that they highlight the basic differences between the two experimental strategies described above. They are ideal criteria in the sense that they would provide valuable guides to the selection of appropriate experiments on the basis of scientific value if they could in fact be applied.

These six "selection criteria" are:

2.3 SELECTION CRITERIA

2.3.1 Minimum Unsupported Assumptions

Every experiment is designed so as to be consistent with certain facts or assumptions regarding the phenomena to be observed. These assumptions are implicit in the experiments in that they provide the informational background in terms of which the results are interpreted. The results of the experiment are not determined by the assumptions, but the design of the experiment, and the interpretations of the results are.

For example, to conduct an atmospheric analysis implies that the experimenter either knows or is assuming that an atmosphere exists. These assumptions are well supported by available observations and theory. Those implicit in other experiments may be weakly supported, or not supported at all in the sense that there may be no relevant facts or theories available.

Assumptions which are unsupported may be more or less restrictive in that the phenomena to be observed may be highly specific. For example, an experiment designed to detect DNA will detect Martian life only under the assumption that Martian life contains DNA. This assumption is less restrictive than the assumption that Martian life not only contains DNA, but is also monocellular, and is characterized by a temperature optimum in the range of 10 to 20°C. On the other hand, an assumption that Martian life-forms exchange CO_2 with the atmosphere and contain polynucleotides of high molecular weight is less restrictive than the assumption that they contain DNA.

One experiment satisifes the criterion of minimum assumptions better than another if its assumptions are either more strongly warranted by available facts or theories, or are less restrictive than those of the other.

This criterion is an important one because experiments which do not satisfy it are, in general, less likely to return useable information for one or both of the following reasons:

1. If the assumptions are supported by available facts or theories we can conclude that the phenomena to be observed have some probability of occurring on Mars. Hence, there is some likelihood that significant data will be obtained. The more strongly the assumptions are supported, the more likely it is that useful measurements can be made. For example, the assumption that organic matter occurs on Mars is more strongly supported than the assumption that proteins occur. Hence, available facts and theories permit us to infer that positive organic matter detection is probably more likely than protein detection and certainly not less likely.
2. If the assumptions are unsupported or only weakly supported, failure to detect the phenomena to be observed does not contribute to our knowledge of Mars because there is usually no way of knowing which of the assumptions are not valid. Geocentric life-detection experiments designed to look for terrestrial-type organisms on Mars are a good example. These return useful information only if the corresponding life-forms are observed. Failure to detect them does not permit us to conclude that they do not exist on Mars.

2.3.2 Criticalness

A fundamental concept employed in the development of the scientific selection criteria is that of the critical question. A critical question is one the program is designed to answer, and there are as many critical questions as there are distinguishable program objectives. The results of an experiment are critical to the extent that they provide information about the existence or nature of life on Mars. Thus, any experiment which has a potential for providing conclusive evidence of Martian life is very critical; experiments which cannot completely answer a critical question are critical to the extent that they support reasonable inferences regarding the probable existence of Martian life or its nature. Thus, criticalness corresponds in many respects to what has been termed "biological interest" or "degree of biological significance" and it is a function of the kinds of inferences which might be drawn from observational data. It is important to recognize that criticalness is not a function of the liklihood that critical results will be obtained, but rather a matter of the degree of significance which might be attached to those results if they were obtained.

Although the relative criticalness of an observation or set of observations is the most important and frequently cited desideratum in judging scientific value, it is almost impossible to estimate by objective means which of two conclusive life-detection experiments is the more critical, or more important, to say how critical observations of environmental parameters might be. In order to estimate in advance of performing the environmental observations how significant the inferences drawn from them would be, it is necessary to be able to list all the possible self-consistent sets of observations which might be made and to identify the inferences which would be drawn from each. This clearly cannot be done. But it does not follow that if a set of observations were available to biologists they are unlikely to be able to draw significant and important conclusions from them.

The major advantage claimed by advocates of the direct strategy is the high degree of criticalness of the conclusive life-detection experiments. The major weakness of arguments in favor of the indirect strategy is the fact that its advocates cannot prove that environmental observations will increase knowledge of the nature of likelihood of Martian life.

2.3.3 Planning Value

One experiment may have planning value for another if, as a result of conducting the first experiment, we are in a better position to judge the value or relevance of the second, or can design a better version of it. For example, a life-detection experiment may have implicit in it a great many assumptions about the nature of Martian life and the Martian environment and to this extent may seem a poor choice for inclusion in an early payload. Another experiment which can test some of these restrictive assumptions will obviously have planning value for the life detection experiment. Planning value is obviously a function of the kind of inferences which can be drawn from the results of the experiment and is thus similar to criticalness. A good example of an experiment which is not very critical but has high planning value, is that designed to measure background radiation on the surface, because it aids in the design of experiments using radioactive tracers.

Experiments which provide information regarding the likelihood that terrestrial organisms will survive on Mars have planning value because they provide data upon which a satisfactory payload sterilization policy can be based. The indirect strategy places emphasis on the early performance of experiments with high planning value, a primary objective of which is to identify which kinds of conclusive life-detection experiments are most likely to provide useful results. A disadvantage of the direct strategy is that failure to observe specific life properties would have little planning value in that they would add little to our knowledge of Mars.

2.3.4 Complementarity

The complementarity criterion applies to sets of experiments which may be included in the same payload or in successive payloads. Two or more experiments are complementary to the extent that their results support one another. Experiments may be complementary in at least two important and distinct ways. Two experiments designed to estimate the value of the same parameter by different and independent procedures are complementary; consistent results permit one to conclude that the estimate is reliable whereas inconsistent results indicate that one is probably a false positive or negative. Two or more experiments which test different parameters are complementary to the extent that the results, considered together, support inferences that could not be as completely supported by the results of each considered independently. It is obvious that experiments cannot be complementary unless there are adequate reaons for supposing that the parameters observed by each of them are related, or that these are jointly related, both theoretically and empirically, to other parameters about which inferences are drawn.

Advocates of the indirect strategy point out that any observations of essential biochemical constituents, growth or reproduction, metabolism or morphology must ultimately be supported by complementary observations of environmental parameters which are consistent with the interpretations of the results of the life-detection experiments.

2.3.5 Reliability

The experiments should be designed to make the inferences drawn from the obsrved samples as reliable as possible. This means that the experiments should be controlled in order to rule out false positives or false negatives due to unanticipated interfering substances. Bias due to instrument malfunction or improper calibration is a source of unrealiability which can be partially controlled by monitoring instrument function or testing with known materials. Random errors due to noise in the instrument, data processing, or telemetry systems are controlled by providing enough observations. Adequate communication bandwidth must be provided. It will be observed that experiments involving few assumptions can be more readily made reliable than those involving many; and that complementarity in the first sense described above leads to enhanced reliability.

It is obvious that an experiment which is not reliable cannot be said to be highly critical in that its data do not clearly support any inference at all.

The reliability of most of the experiments cannot be estimated at this time because flight prototypes are not available and because data processing and data communication systems have not been designed yet.

2.3.6 Contamination

All experiments and payloads must be sterilizable to the degree required to satisfy policies regarding acceptable levels of contamination risk. The extent to which experimental instruments can be made to satisfy this criterion cannot be evaluated now.

2.4 Review of Criteria

It will be noted that two of the criteria listed above may be said to correspond to minimum necessary -- but not sufficient -- conditions which any experiment must satisfy before it can be considered a candidate for inclusion in a payload. These are the sterilization criterion, which relates to the extent to which experimental instruments satisfy the sterilization policy, and the reliability criterion, which concerns the probable reliability of the various possible inferences to which the experimental results might give rise. Two criteria -- those of minimum unwarranted assumptions and planning value -- consider the likelihood that an experiment will provide data that will support specific useful inferences. The criticalness criterion considers the differential value of those inferences in drawing important conclusions about Martian biology or in planning subsequent manned or unmanned explorations. The remaining criterion -- that of complementarity -- concerns ways in which several experiments in a single payload or in successive payloads may fit together so as to provide mutually supporting observations or to constitute partial checks on each other.

Although these criteria usefully served to focus attention on what appear to be real and significant differences between the two experimental strategies, it did not provide a practical guide to the selection of a strategy or a payload because it proved very difficult to actually evaluate all individual experiments in terms of them, partly because some of the experiments were insufficiently developed, and partly because some of the evaluations seemed to require the use of subjective estimates, on which little agreement could be expected. An additional difficulty is due to the fact that the criteria are in fact not separate, but complexely interrelated.

An informal statistical decision model of the payload selection problem was therefore developed in order to provide a less ambiguous interpretation of these criteria, to explicate some of their interrelations, and, most important, to aid in sorting out those kinds of judgements which can now be made about alternative experiments from those which cannot be made at this time, and which should therefore best be dropped from consideration.

The model which was developed is an informal discursive one. That is, it consists primarily of a set of definitions in words of basic concepts, and its conclusions consist mostly of inferences developed from these definitions by means of closely reasoned arguments rather than of theorems derived from a set of formally expressed postulates. The primary relevant conclusions which appear to be supported by this model are the following:

2.4.1 1. The most useful criterion is that of reliability. Its utility derives from the fact that it is widely applicable and the fact that it specifies a minimum condition which ought to be satisfied by all candidate experiments.

2.4.2 2. Relative freedom from unsupported assumptions is positively correlated with reliability. As a consequence, it makes little sense to speak of the criticalness of experiments which do not satisfy this assumptions criterion because such experiments, on account of their unreliability, cannot be guaranteed to provide results which support any inferences at all.

2.4.3 3. Biology is "unconnected" in the special sense that it is rarely possible to infer, from the knowledge that some biological properties exist, whether a great many others are more or less likely to co-exist with them. Although no scientific discipline actually possesses the character of a hypothetico-deductive system in which all conclusions are arrived at in a fashion analogous to the derivation of a theorem in a highly formalized mathematical system, some of them have a distinctly more formal character than others. Physics and chemistry are frequently cited examples of relatively formal disciplines and in this respect may be contrasted with physiology and psychology. In the more formal and mathematical disciplines there tends to exist a relatively codified body of accepted theory and implicit translation rules which relate to basic theoretical concepts to observable phenomena. There is as a result somewhat greater agreement among practitioners as to which conclusions follow or would follow from which observations, and consequently it tends to be easier to specifiy in advance those kinds of observations which are of greatest importance in the acceptance or rejection of specifiable hypotheses. In saying that biology is "unconnected" we mean merely that it is relatively non-formal. An important consequence of this truism is that it is generally not possible to speak other than metaphorically of the "amount of information" represented by a given observation because one can specify neither the "possibility space" of alternative Martian biologies, nor the different inferences to which the alternative possible results of a given experiment might give rise. As a consequence, it is generally not possible to specify in advance on objective grounds the extent to which specifically biological*experiments satisfy the planning value and criticalness criteria.

Estimates of the planning value and criticalness of such experiments must be subjective, and it can be expected that there will be widespread uniformity of judgement only with respect to those experiments judged to be of only moderate criticalness (e.g., temperature measurements). This being the case, the failure of a given experiment to be considered highly critical does not constitute good grounds for excluding it.

A parallel consequence of the unconnectedness of biology as a scientific discipline is that the degree to which experiments of a biological nature satisfy the minimum unwarranted assumptions criterion can be estimated only crudely. The only experiments which satisfy this criterion to a high degree are those which measure physical properties on a scale which includes a zero point, for such experiments are bound to provide some information about the state of Mars and to exclude a great many possible other states. (A good example of such an experiment is an atmospheric sampler designed to determine the relative quantities of each of a number of gaseous constituents: each such constituent is either present in some measureable concentration or absent; from the observation that

*The distinction is between experiments involving observations of biological phenomena and experiments involving measurements of physical or other non-biological properties which support conclusions about biology.

one mixture is present it is possible to infer the absence of all others).

2.5 Decision Model of the Experiment Selection Problem

The decision model developed during the course of the summer study is briefly presented in this section in non-technical terms. A more formal and more sophisticated model which was also employed in the analysis of experiment selection problems conducted during the course of the reporting period is being prepared for independent publication.

2.5.1 Primitive Notions

Any program of biological exploration of Mars may be viewed as consisting of a sequence of inquiries corresponding to the biological payloads launched toward the target planet at each opportunity. Each payload is an instrument designed to conduct one or more experiments and to return to earth information regarding the consequences of that experiment. It is useful to construe each such experiment (or, more precisely, the particular instrument designed to conduct that experiment) as the physical representation of a question regarding the presence on Mars of a particular property or member of a set of properties.

Thus, one proposed metabolic experiment can be said to determine whether, at the time and place of impact, there exists in the sample collected a sufficient quantity of organisms capable of metabolizing and fixing a sensible quantity of a detectable metabolic product. In making such a determination the instrument which performs this experiment asks whether Mars has the property of causing it (the instrument) to respond in the specified way. Since Mars may have this property without its being observable at the time and place of impact, we may say that the instruments are in principle capable of demonstrating the presence of properties on Mars, but not their absence. (Those few exceptional cases where the absence of one property can be inferred from the presence of another are considered separately below in the discussion of the exclusive property detector).

It will be noted that conceiving of each experimental instrument as an inquiry regarding the presence on Mars of a single property is a gross over-simplification of the real situation, in that it implies that each instrument is capable of providing only one bit of information (property detected or not detected). Actually, each instrument is more properly conceived of as estimating the value of one or more parameters, and the number of possible parameter values that could be distinguised by the various detectors is generally quite large.

We may argue that the simplified view of each instrument as detecting a binary property is not overly unrealistic by pointing out that the ranges of possible values of the n - dimensional parameters observed are often spoken of as divisible into two classes - those which are and those which are not biologically significant, since many properties, if they are present at all, are present in biologically significant quantities. Furthermore, although the sensitivity and dynamic range of the sensors may make them capable of distinguishing a large number of parameter values, sampling error, telemetry limitations, instrument bias, and the problem of interference from unanticipated materials yielding false positives or false negatives will remain very real possibilities with even the best designed experiments. Consequently, inferences based on the observations will be far less precise than the sensitivity of the sensors would suggest, and we may safely generalize that the number of parameter values which can confidently be distinguished is not nearly as large as might first be supposed.

There is, however, one set of detectors that cannot be so readily viewed as binary property detectors. These are those detectors which, like an atmospheric gas sampler, always give a "yes" answer in that they always detect the presence of something on Mars. We may examine the problem represented by this class of detectors by considering the atmospheric sampler in more detail. The number of different mixtures which can be distinguished is very, very large, even when allowance is made for various kinds of errors. Since we can in general infer from the observation of any one of these mixtures the absence of all of the others, it is not intuitively satisfactory to conceive of such an instrument as being a binary property detector. We can accommodate these reasonable reservations by explicating the special notions of a <u>set of exclusive properties</u> and an <u>exclusive property detector</u>. This requires dividing the set of possible mixtures into a fairly large number k of subsets $M_1, M_2, \ldots M_k$ each of which consists of a family of possible gas mixtures such that all of the mixtures in that family are similar with respect to the biological significance and physical implications of the inferences which might be drawn from the observation of any one of the mixtures in the family. Each of the k subsets then corresponds to a property. These properties $M_1, M_2, \ldots, M_k$ are mutually exclusive in that there is only one degree of freedom: exactly one of the k properties will be detected and the observation of one permits the inference that all the others are absent in the atmosphere in the immediate vicinity of the lander. Thus an exclusive property detector may be thought of as a set of K binary property detectors where the sum of the probability of occurrence of each property ($pM_1 + pM_2 + \ldots + pM_k$) is 1.

At each opportunity one may ask questions about the presence of properties on Mars, and the problem is to decide which are the best questions to ask at which opportunities, assuming of course, that there are or could be more instruments available to be sent than can be sent at any single opportunity. Since in general only yes answers constitute useful information, the decisions as to which questions to ask is partly a function of the probability that a yes answer will be received and partly a function of the relative value of knowledge about the presence of the different properties (in that, e.g. some properties are more decisive indices of the presence or absence of life than are others.) Let us ignore for the time being the question of the relative value of knowledge of different properties and consider the factors that influence the probability that asking a given question will result in a yes answer.

The "question" is an instrument which, in order to return an answer of any kind, must be successfully transported to Mars, land, and there survive in satisfactory working order long enough to conduct the experiment and transmit the results back to earth. A "question" which does this may be said to be an <u>engineering success</u> and for each question and each launch opportunity, there exists some P_E in the range 0 to 1 which is the probability that the probe will be an engineering success. We shall use Pe_i to denote the probability of engineering success of the instrument which detects the property i.

If the probe is an engineering success, it will return a "yes" answer only if Mars has the property in question and if it is detectable at the time and place that the experiment is conducted. Thus for each question there is some probability P_{M_i} that the corresponding property i exists on Mars in the requisite time and place such that if the probe is an engineering success a yes answer will result.

The notion of the probability P_{M_i} of occurrence of some property i on Mars needs some explication. Ideally, one would arrive at an estimate of this probability by enumerating all the possible states $s_1, s_2, \ldots s_n$ of Mars that are consistent with our present knowledge of that planet. Some of the states in the resulting set S will contain P_i and some will not. If the property P_i is homogeneously distributed in space and time about the planet, then P_{M_i} is the relative frequency of states containing P_i in the set S of all states of Mars. If the property i is not homogeneously distributed, then one must estimate the likelihood of encountering P_i in the locations and at the times the instruments are likely to land. P_{M_i} is then the product of the relative frequency of states containing P_i in the set S and the "encounter probability" of P_i in each state. Since it is impractical if not impossible to enumerate all the conceivable states of Mars, P_{M_i} cannot be readily estimated. The set of properties observed by the exclusive property detector is a partial exception to this rule in that the probability that <u>some</u> one of the k properties will be detected is always 1, although it will often be difficult to enumerate K subsets and assign appropriate P_Ms to each. There are, in addition, other properties about which we already have some direct or indirect evidence, for which moderately accurate and precise estimates of P_M can be generated.

Although the notion of P_{M_i} as relative frequency of occurrence of states containing i in the set S of all states of Mars (modified, if necessary, by estimates of encounter probability) is not one which can in practice always be evaluated, it is nevertheless useful in that it can be seen to correspond to the notion of "freedom from unwarranted assumptions". That is, the hypothesis that a property is present is free from unwarranted assumptions to the extent that P_M is high. Thus experiments designed to determine which of a set of exclusive properties occurs on Mars always satisfy this criterion.

In order to apply a decision theoretic model to decision problems in extraterrestrial biology we must be able to interpret the notion of the <u>utility U</u> associated with the detection of each property or set of properties on Mars. Intuitively speaking, the utility of an outcome of some act which could be chosen by a decision maker is the advantage accruing to him of realizing that outcome.*

The purpose of estimating utilities is to permit a comparison of different acts in terms of the relative advantages associated with their outcomes. Consequently, it is desirable to express relative advantage in terms of a position on an interval scale. It must make sense to say, for example, that the utility associated with detecting property x is half that

*The notion of utility presented here departs somewhat from conventional notions of utility in that we distinguish the advantages (including negative ones) of the possible outcomes of an act from the costs of executing that act. We do this because we wish to maintain a distinction between dollar and other costs and less tangible utilities, in particular those corresponding to scientific value.

associated with detecting property y and equal to that associated with detecting property z and that, as a consequence, detecting y is equivalent in utility to discovering both x and z.

Although the concept of utility will not be defined here, we may provisionally suppose that it corresponds to what might be termed "relative scientific value of experimental results", and as such is not unrelated to the concept of criticalness explicated above. Two properties may be said to differ in criticalness to the extent that evidence of their occurrence on Mars permits one to draw conclusions regarding the past, present, or future occurrence of life on that planet. If it is possible to rank experiments according to their relative criticalness, (and the use of this concept in arguing about experiments suggests that this ought to be the case) then it should also be possible to rank them according to their relative utilities.

We shall assume provisionally that this is the case, and also make the slightly stronger assumption that utility is measurable on an interval scale. That is, we assume that we can assign to each property a position of a utile scale of e.g. 0 to 100 or 1000 <u>utiles</u>. It should be pointed out that assuming that scaling of criticalness does not require us to suppose that discovering the presence of two properties, for example glucose and organic molecules of molecular weight in excess of 5000, to which we might have made utility assignments of 40 and 30 respectively, is the <u>same as</u> discovering net optical activity, to which we might have assigned the utility value 70, in the sense that having discovered the presence of the first two properties we no longer need to look for the third.

To say that the utilities associated with two subsets of properties are equal does not mean that to know that one occurs on Mars provides us with any information regarding the occurrence of the other. (This kind of value receives another definition below). We also do not mean that the utility (or criticalness) of an experiment is not a function of our knowledge. In general, our estimates of the criticalness of an experiment will change with increases in knowledge of Mars or of terrestrial biology. This means that the utility we assign to discovering that Mars has the property of causing a particular instrument to respond in a specific way at one opportunity may not be the same as that assigned to it at a subsequent opportunity. The increase in our knowledge of Mars resulting from the experiments conducted during the first opportunity may cause us to revise our estimates of the utility associated with the discovery of that property or some other at the next opportunity. It will be noted that utility may change with changes in the state of knowledge independently of shifts in estimates of the probability of occurrence on Mars of the property*.

---

*For example, net optical activity is believed to be co-extensive with terrestrial life in that all biogenic matter exhibits this property and no non-biogenic material does. It has also been suggested that this is a necessary property of life, but the latter view is not universally held. If, in the next few years persuasive evidence to the effect that some biogenic matter is racemic and hence net optical activity is not co-extensive with life becomes available, then the utility of finding this property on Mars will be reduced. On the other hand, the utility of finding this property on Mars could be increased by the development and acceptance of a better theoretical basis for supposing it to be a necessary property of life.

The concept of criticalness and its correspondence to utility will be more carefully reviewed in a sebsequent section, in which we introduce additional notions of utility. For the present we shall assume that the concept of criticalness is sufficiently valid to permit us to assign to each experimental instrument some utility value on a scale of 0 to 100 which reflects the best estimate, based on our current knowledge, of the relative criticalness of each of the corresponding properties. We may then define the theoretical expected value (abbreviated TEV) of the probe consisting of the ith instrument as the product of P_{M_i} and U_i, and we may define the practical expected value (PEV) as the product of the theoretical expected value and the probability of engineering success.

$$TEV_i = P_{M_i} U_i$$

$$PEV_i = P_{M_i} U_i P_{E_i}$$

We distinguish between these two types of expected value primarily because it has been suggested that decisions regarding the scientific value of alternative strategies of inquiry be made, initially at least, independently of any consideration of factors associated with reliability and cost, or of possible "embarrassment factors" such as looking for something that is later shown not to exist, or launching a series of engineering failures. However, it is obvious that if estimates of the reliability and cost of the alternative instrument packages were available then choices could be made on the basis of practical expected value, and that useful cost comparisons could bemade in terms of practical expected value per unit cost (where the cost need not necessarily be a function of dollars spent but may reflect probability of contamination or any other relevant parameter which a decision-maker may wish to minimize).

An illustration of the manner in which theoretical and practical expected value may be used to rank selection is shown in tables 1 and 2. Assumed precise and accurate estimates of P_m, U, P_E and dollar cost are tabulated for five properties together with the corresponding theoretical and practical expected values and a calculated practical utiles/unit cost (Put/\$). Table 2 shows the different orderings which in this example result from the three different choice criteria. In this example the range of costs and P_E values is not great, and consequently the three orderings do not differ as radically as they might in a more realistic example.

It will be observed that if one expands the set of alternatives to include probes consisting of 1, 2, ... n of the set of n available different property detectors, the best choice will consist of a package carrying all n detectors if theoretical expected value is the only criterion employed in making the choice. If, however, all have the same probability of engineering success and cost the same, then the best payload from the point of view of practical expected value per unit cost will consist of the single instrument with the highest theoretical expected value. This is illustrated in table 3, which compares three payloads composed of one or more of the first three properties listed in table 1. This example makes the unrealistic assumption that engineering reliability is not reduced by integrating the payload, and that the utility of knowing that three properties co-exist in the vicinity of the lander is not greater than the utility of knowing the same three exist at different times and places on the planet.

TABLE 1

Property	P_M	U	P_E	$	TEV	PEV	Put/$
1	.06	70	.7	30	4.2	2.94	.096
2	.01	90	.7	30	.9	.62	.02
3	.20	16	.7	30	3.2	2.24	.075
4	.70	5	.9	35	3.5	3.15	.093
5	.2	18	.8	40	3.6	2.88	.072

TABLE 2

Criterion					
TEV		PEV		Put/$	
Property	Value	Property	Value	Property	Value
1	4.2	4	3.15	1	.096
5	3.6	1	2.94	4	.093
4	3.5	5	2.88	3	.075
3	3.2	3	2.24	5	.072
2	.9	2	.62	2	.02

TABLE 3

Components	Σ TEV	Σ PEV	Σ $	Σ PEV/Σ $
1	4.2	2.94	30	.096
1 & 2	7.4	5.18	60	.086
1, 2 & 3	8.3	5.80	90	.064

2.5.2 Plans

We have proposed that the decision maker act in such a way as to maximize his expected value. He will decide which acts to implement on the basis of estimates of P_{M} and U, if he employs the choice criterion of theoretical expected value, and will also employ estimates of P_E and of dollar or other costs if he uses either practical expected value or value per unit cost as a criterion. We shall say that P_M, P_E, U, dollar cost, etc., are planning parameters and that estimates of their values constitute a set of planning information items. Given these estimates the decision maker can rank the possible acts in order of decreasing expected value, of which he will select one or more of the top ranking acts. The plan to which such a ranking gives rise is the act actually selected, and if the decision maker adheres strictly to his choice criterion, the plan will consist of the top ranking item. However, since not all factors actually relevant to the decision are formally taken into consideration by the choice criterion, the decision maker normally employs the criterion to identify a subset of top-ranking alternatives and then selects one from this set on the basis of additional relevant considerations which are not explicitly taken into account by the criterion (usually because they cannot be quantitatively expressed). It is, therefore, somewhat more realistic to conceive of the plan derived from a criterion and a set of planning information items in terms of some subset of high ranking items, and to consider the differences between criteria and/or sets of planning information items in terms of the differences in the identities of the elements in the subset, the magnitudes of the expected values assigned them, and the rank ordering defined over them by these differences.

Table 4 shows four assignments of expected value and the resulting orderings defined over the same six acts. Considered from the point of view of the identify of the top ranking three acts, orderings one, two and four are identical in that all show a, b, and c to be the highest ranking trio. Considered from the point of view of the actual order, however, ordering two differs from one and three in that the positions of a and b are reversed. These three orderings of all six acts differ relatively little in that the positions of no act differs by more than one rank, and in this respect ordering three is seen to be radically different from the others.

Two plans which are similar in respect to the positions assigned acts may differ greatly in terms of the differences in value assignments made to the individual acts. We shall say an ordering is selective to the extent that these differences are large in absolute value or in proportion to the total range of values represented.* Thus, ordering three is a strict ordering in that no two acts are assigned the same value, but the absolute differences are so slight as to lead the decision maker to infer that the reduction in expected value resulting from choosing the second or the third ranked act rather than the first is so slight as to make them almost a toss up. The same is true of ordering four, which shows two ties (b and c for

*Selectivity defined in terms of absolute value is appropriate only where we have reason to believe the value units employed in the different orderings are comparable. If the scales must be normalized by a linear transformation, then selectivity is a function of proportional difference.

second place and d and e for fourth) and a difference of only three points out of a range of 72 between the first and second positions. Selectivity is important because in general the greater the differences between assigned values, the less likely it is that slight errors in the planning information will result, when corrected, in great differences in rank order, which is another way of saying that the choice is more obvious.

One plan is better than another to the extent that it permits the decision maker to maximize his expected utility, and a plan which defined the correct ordering but is not selective in that the value assignments are incorrect can be just as misleading as one which shows the wrong order but has more accurate value assignments. For example, if the value assignments in ordering one were the correct ones, then the use of ordering four might lead the decision maker to select act b or act c on the grounds that the loss of three points of expected value is a very slight penalty in comparison with other advantages he believes to be associated with these. As a result he may suffer a lost of 20 or 40 units of expected value. If he employs ordering two he is unlikely to choose c in spite of its advantages and, hence, will suffer a loss of not more than 20 units of expected value.

TABLE 4

Ordering 1		Ordering 2		Ordering 3		Ordering 4	
act	value,	act	value	act	value	act	value
a	80	b	70	d	23	a	81
b	60	a	66	f	22	b	78
c	40	c	30	c	19	c	78
d	10	e	15	a	16	d	60
e	5	d	10	b	12	e	60
f	1	f	8	e	9	f	10

2.5.3 Informational Dependence

Considerations of the scientific value of alternative experiments are usually focused on questions of the kinds of inferences to which the experimental data may give rise--that is, on how a given set of observations may be interpreted so as to yield conclusions regarding the state of Mars encountered by the experimental equipment.

In considering questions of evaluating alternative experiments as a function of the inferences which may be drawn from their results, it is important to distinguish carefully between at least two interpretation steps. The first interpretation step is that whereby the messages transmitted from the planet are interpreted as describing the state of the experimental equipment and the nature of the responses of the sensing units included in that equipment. This interpretation step requires diagnosing the nature of the response of the on-site measuring equipment and determining whether this response accurately reflects the Martian conditions encountered, or is due to any of a number of possible sources of error. (These problems associated with various kinds of reliability are more completely discussed in sections 3, 4 and 5 of this report.) If the experiment is designed so as to be reliable in the sense that it is possible to distinguish responses which may be confidently attributed to properties the experiment was designed to observe from those attributable to the various sources of error, then a second interpretation step can be performed. This step consists in concluding from the presence or absence of the properties the experiment was designed to observe, the probable presence or absence of other properties of biological relevance.

It will be noted that if an experiment is not sufficiently reliable as to permit the first interpretation step to be effected with some confidence, then the second interpretation step is not possible. In the following discussions we assume that all experiments are reliable in this sense.

If observations of the presence of one property at one location and one time provide knowledge about the probable presence of other properties (at the same or other locations and times), then the results of the first experiment can be said to support inferences regarding the other properties.

The process of inference is one whereby knowledge of the presence (and sometimes the absence) of one property permits one to modify estimates of the probability of occurrence of another property. We shall say that two properties (or two subsets of properties*) a and b are informationally dependent when the estimate of P_{m_a} given knowledge of the existence of b is different from the estimate of P_{m_a} given uncertainty of the existence of b. We may symbolize this relationship of informational dependence as

$$P_{m_a}/b = P_{m_a}/?b$$

where "P_{m_a}/b" should be read as "estimate of the probability of occurrence of a on Mars given knowledge of the existence of b" and "/?b" as "given uncertainty of the existence of b". If we further define "/-b" as "given knowledge that b does not exist," then the kind of informational dependence that exists between two mutually exclusive properties may be expressed as

$$P_{m_a}/-b = 1 \text{ and } P_{m_a}/b = 0 \qquad P_{m_a}/?b + P_{m_a}/?a = 1$$

If the presence of a is a necessary condition for the presence of b, then usually

$$P_{m_a}/b \text{ greater than } P_{m_a}/?b$$

*For example, sets of values of a number of parameters.

and if it is also a sufficient condition, then

$$P_{m_a}/b = 1 = P_{m_b}/a$$

We may say that a set of properties if highly connected if informational dependence relationships exist between many of its members. The amount of connectedness in a set of properties is a function of our knowledge of correlations between pairs of its elements. In general, correlations of which we have no explicit knowledge or which cannot satisfactorily be explained on theoretical grounds do not results in informational dependence.*1

It will be noted that if informational dependence relationships do exist, it will in general be possible to estimate the conditional probabilities of occurrence of various combinations of properties in a connected set. The less connected the set, the fewer the combinations to which conditional or joint probabilities of occurrence can be assigned. In general, the inability to make some estimates of joint probabilities of occurrence reflects the fact that observing the properties in the set will not give rise to inferences.*2

*1The distinction between what we do infer from available data and what we ought to infer or would infer if understanding of the phenomena were more perfect is an important one which must be kept in mind in any discussion of the amount or the value of information available from an experiment. This notion of connectedness will be seen to be very similar to that of degrees of freedom. The term "connectedness" is employed here because the reference is not only to those situations where the connectedness relationships are supported by widely accepted theoretical considerations or empirical evidence, but also those which can be said to describe shifts in subjective estimates of likelihood on the part of qualified scientists. That is, we wish to include the consequences of judgment based on scientific intuition and hunch.

*2It is important to note that we are not concerned here with the basis of such estimates. They may be purely subjective. We merely note that failure to be able to make such estimates (even subjective ones) tends to predict failure to be able to draw inferences once the properties have been observed.

If there is some theory which makes it possible to determine deductively (e.g., through calculation) that certain combinations are impossible, then those properties will be highly connected in the sense described here. Exclusive sets are sets of properties which are highly connected in this fashion, even though some such sets will be trivially connected in the sense that it is not logically possible for one parameter to have two values simultaneously. (The difference between connectedness arising from purely logical considerations and that arising from the fact that scientific theories are sufficiently well established to make some assertions regarding the impossibility of some combinations tautologous is not significant, and it can be maintained that none of these informational dependence relationships are truly trivial.)

The higher the estimated a priori probability of occurrence of a property as estimated on the basis of available knowledge of Mars, the more strongly the corresponding properties are connected to properties of which we already have some knowledge. It follows that experiments which satisfy the minimum unwarranted assumptions criterion can, in general, be guaranteed not only to provide some information (because a successful experiment is more likely to result) but will also tend to provide data on which inferences can be based.

The concept of informational dependence can be employed as a basis for developing notions of the inference value of a successful experiment. Given a set of properties among which some informational dependence relationships exist, it is possible in theory to generate the set of all logically possible combinations of the presence and absence of those properties corresponding to all logically possible combinations of results of the corresponding experiments as they might be performed simultaneously at one site or at different sites on the planet Mars. These combinations may then be divided into the following subsets:

1) Those which represent physically or logically impossible combinations, or combinations which are so exceedingly unlikely as to be not worth considering.

2) Those which are physically possible but unlikely.

3) Those which are viewed as probable.

It will be noted that if the experiment is site and/or time dependent such as is the case with experiments which are "successful" only if they result in positive detection of the corresponding property, then class c will be very, very large. If on the other hand the property to be observed is homogenously distributed over the surface or through time, then class c will be small.

The inference value of a single experimental result can be estimated by considering the amount by which class c would be reduced if that result were observed. If the set is highly connected, then knowledge of the presence or absence of one of the properties will make certain otherwise likely combinations unlikely or impossible. It is obvious that this measure of inference value can be extended to include the possible outcomes of sets of experiments. The amount of information provided by a given experimental result can then be defined as a linear function of the inference value and the a priori probability of occurrence of that result. If a particular result is almost certain, then that experiment will provide less information than would be the case if the result were less likely. If the inference value of a given result is very low (as is the case with a negative answer from a highly site dependent experiment), then the amount of information provided by it will also be very low no matter how likely or unlikely that result. A positive result from a highly site dependent experiment will also be seen to provide relatively little information in the sense that its inference value will also be low.

It will be noted that these notions of inference value and relative information content are all defined relative to a set of connected properties and estimates of the joint probabilities of occurrence of members of that set.

At this point it should be pointed out that the reliability of the inferences drawn during the second interpretation step described above is a function of the validity of the assumed informational dependence relations supposed to exist between properties in the connected set. If these relationships are based on subjective estimates (and if they differ with the identify of the scientist making them), then the inferences cannot be demonstrated to be reliable. Thus, if we demand that an experiment be a reliable one then we must restrict our consideration to experiments involving sets of connected properties whose informational dependence relationships can be validated in a conventional way, or those where there is wide agreement regarding the subjectively-arrived-at estimates of informational dependence.

2.5.4 Utility and Scientific Value

The preceeding discussion of informational dependence and the kinds of inferences to which experiments may give rise demonstrates the complex nature of the interdependence between the criteria presented above. If this concept is a valid one then it follows that the minimum unwarranted assumptions criterion and the reliability criterion can be defined so as to provide a basis for estimating the inference value and relative information content of alternative possible experimental results. It is proposed that the notion of utility in the sense of relative scientific value can be usefully based on such estimates. One of the consequences of this model of scientific value is the fact that the "biologically interesting" experiments of alleged high criticalness turn out to have little or no scientific value. This is due to the fact that such properties tend to have one or more of the following characteristics:

1) They are highly site-dependent and consequently negative results have little inference value (positive results may also have little inference value for the same reason).

2) The estimates of informational dependence are highly subjective (because the lack of well developed theories in biology results in there being few "objective" grounds for estimating informational dependence) and tend to vary greatly among estimators. Consequently, reliability of inferences cannot be readily demonstrated on objective grounds.

3) These experiments cannot readily be made reliable insofar as the first interpretation step is concerned; it is difficult to distinguish valid positive results from false positive results, and consequently

these experiments cannot readily be said in advance to be sure to provide any "results" at all.

It was pointed out above that other kinds of utility can and should be considered in planning an exploration program. The use of the choice criterion of practical expected utility per unit cost in dollars implies a notion of utility based in part on dollar cost. Some other utilities which could be employed in addition to or instead of scientific information content are:

1) Embarrassment cost of engineering failure. A single failure could be supposed to entail a unit negative value. The negative value of a sequence of failures could be supposed to increase exponentially with the number of failures. The unit cost per failure might be supposed to be a function of the dollar and time cost associated with developing and launching the system.

2) Embarrassment cost of looking for something that does not exist, or failing to look for something that does. If a careful attempt has been made to estimate scientific value as defined above in terms of inference value and relative information content, then there will be little or no embarrassment value of this kind. However, such estimates are of necessity intuitive because the model proposed here for didactic purposes cannot easily be implemented by practical procedures for evaluating these factors. As a result, it is always possible that after a payload has been designed and/or launched, new evidence or new information will cause revisions of the utility estimates employed. This is less likely to occur if the choice criterion employed in generating the plan is highly selective.

3.0 RELIABILITY AND ERROR CONTROL

In section 1 it is noted that among the several criteria entering into the selection of an experiment or mix of experiments for the biological exploration of Mars, one of the most heavily weighted is that of reliability. Certain aspects of the problem of experimental reliability is further discussed and illustrated in section 4 of this report, in which the outputs of some simple stochastic simulations are presented. In particular, in the second of these simulations, SAMPLER, a simple system is examined under varying probabilities of component failure in order to obtain a qualitative picture of the over-all probability of what we have called "mission vitiation." As used in this report this term refers to a situation in which the system fails for one reason or another to exercise suitably "tight" control over the performance of the two halves of a contrasted experiment. Such a situation obtains when the statistical properties of the physical system are mismatched with the physical-biological environment or when actual component failure prevents the performance of one portion of the contrasted experiment. Under such circumstances the investigator may receive accurate data on both halves of the experiment yet be quite unable to confidently interprete his results, or he might be in receipt of a "half experiment" which he is quite unable to appraise.

3.1 Reliability

From the foregoing it will be noted that the term "reliability" is used in this document in a wider sense than that generally employed in the engineering literature. As used in this report, in fact, the term denotes the entire spectrum of physical-engineering acts and events whose concatination will culminate in data which can be interpreted with confidence in terms of whatever hypotheses the investigator is testing. This section is devoted, therefore, to a brief definition of the terminology used in the report.

(1) Engineering Component Reliability (ECR). The "conventional" reliability in that it is the quantity

$$P(t) = P(0) \cdot f(t) \tag{3.1.1}$$

which is the probability of the successful operation of the component in the time interval t to t + dt.

(2) System Topology Reliability (STR). The probability that an entire system will successful function within designed limits during the interval t to t + dt. STR is dependent not only on the individual ECRs of each component, but also on the pattern of interconnections between the components, their sequence of operation and on the crucialness of each component.

(3) Noise. The perturbation of a useful datum signal or variable by a random component e(t) arising from uncontrolled sources. By definition

$$\left.\begin{aligned} E[e(t)] &= \int e(t)dt = 0 \\ E[e^2(t)] &= \sigma^2 \end{aligned}\right\} \tag{3.1.2}$$

so that the long time average of noise is zero. In general, if the experiment looks at a variable $\mu(t)$, the quantity seen on earth is

$$X(t) = \mu(t) + e(t) \tag{3.1.3}$$

and the average of X(t), i.e., AVE X(t), is $\mu(t)$ so that by averaging long enough we can suppress noise effects.

(4) Bias. The perturbation of a useful signal by a component b(t) arising from drift, damage, calibration or detector failure, etc., which does not average to zero. Then, if

$$X(t) = \mu(t) + b(t) + e(t) \tag{3.1.4}$$

is the received signal, AVE $X(t) = \mu(t)$ + AVE $b(t) \neq \mu(t)$, so that the "true" value of the variable $\mu(t)$ cannot be recovered from the data.

(5) Error of the Third Kind. Occurs when bias arises in the data and is not detected by engineering monitoring or by calibration. Then the investigator draws conclusions based on $\mu(t)$ + AVE b(t) which may have nothing to do with the true scientific problem represented by $\mu(t)$.

(6) Confounding and Aliasing. Obtain when a detector responds to two or more quite different types of phenomena as if they were the same. In, for example, a labeled substrate conversion to labeled product by mediation of an organism, some of the label may be released from the medium by abiogenic means and some by biologic mechanisms. Insofar as the detector is concerned the two release mechanisms are aliased or confounded together so that they cannot be distinguished.

(7) Accuracy and Precision. When a source of data is perturbed by noise (3 above) but not by bias (4 above), they are accurate in the sense they tend to average so that noise is suppressed. If, in addition, the noise level is low so that the data are not greatly perturbed, they are also said to be precise. Precise data, on the other hand, may not be accurate due to bias in the readings.

(8) Data Window. The length of time, from deployment to termination, that a device or system is designed to function.

(3) Markovian Versus Non-Markovian Failure Processes. When the P(t) of a particular component is really a function of the number of times the component is operated, we refer to this as a Markovian failure process. Alternatively, if P(t) steadily decreases with passage of time independently of operation, this is referred to as non-Markovian. The mixed case is considered non-Markovian.

3.2 Design of Automatic Experiments

In the performance of terrestrial experiments it is usually, but not always, a relatively simple problem to perform the necessary checks and calibrations to obtain reasonable assurance of unbiased operation of the equipment at tolerable noise levels. Further, adequate duration of operation of the equipment is easily designed on the basis of a few preliminary experiments.

Such is not the case in remote, automatic experiments where all contingencies must be provided for in advance in the design of the device.

In implementing a particular experiment it may be assumed that ECR of each component will be as high as possible. In general, in the early Martian probes, battery life will determine the length of the data window so that the STR, S need not be greater than L, the operating life of the power source. This, however, poises a problem because in general ECR can be expressed as

$$P(t) = P(0) \exp - \int_0^t H(t)\, dt \qquad (3.2.1)$$

where H(t) is known as the hazard function and is a complex function of component fabrication methods and sterilization procedures. Since **P(t)** steadily decreases from its deployment time value of P(0), the question exists of to what levels can P(t) be safely allowed to fall by t=L, the end of the mission, to give reasonable chances of success of the mission. The answer depends partly on the nature of the experiment, partly on the amount of payload which can be devoted to engineering redundancy, and partly, but significantly, on the form of H(t). The latter, however, is largely a matter of money expended to engineer a suitable H(t) for each component and this can in principle be determined by knowing the relative weight the investigator will give to each successive datum and the adequacy of the failure mode monitoring so that he is sure to know when failure has occurred.

Unfortunately, it turns out for reasons to be explained later, that the STR during the terminal stage of an experiment's lifetime is a critical parameter in estimating the probability of an experiment being performed successfully. Before discussing this, however, a brief digression will be made concerning heuristic systems.

3.2.1 Heuristic Systems

It has been proposed that consideration be given to landing a large automated laboratory (ABL) on Mars with the capability of designing and executing its own experiments using the results of prior experiments as design parameters. Since what the laboratory does on each cycle is dependent on what it found on the previous cycles, its equipment and component duty cycles will be random events. Assuming the failure processes will be non-Markovian (mixed), we are confronted with evaluating the STR of the experiment under the condition that the configuration of the system at any time is a random process. This likely to be a difficult problem but must be solved in view of the likely high cost of such heuristic systems.

3.2.2 Importance of Terminal ECR and STR

Consider, as an example, a simple growth experiment to test whether a sample of Martian material will reproduce according to

$$\begin{aligned} \ln N &= \ln N_0 + \beta(t-1) & t > l \\ &= \ln N_0 & L > t > 0 \end{aligned} \qquad (3.2.2)$$

where l is the lag period. Assume the experiment is power limited to a total total of L time units. Let the STR of the system be expressible by

$$\begin{aligned} P(t) &= P(0)\,e^{-\lambda t} & t \leq L \\ &= 0 & t > L \end{aligned} \qquad (3.2.3)$$

and for simplicity, let P(0) = 1. Since we have no idea what value l will really assume (it might be greater than L), we are forced to postulate that it is probably L/2 since we have no basis for any other value. Similarly, we have no real knowledge of λ so that we might as well assume that it is at least $2\Delta/L$ where Δ is the smallest difference in two responses that can be discriminated on the telemetered output. Finally assume that failure mode monitoring will alert the investigator if failure occurs.

The amount of data collected will obviously be related to the rate at which P(t) decreases. Assume the worst case $l = L/2, \beta = 2\Delta/L$.

In order to obtain N barely usable data points the system must remain operative through the time period L/2 to L. The actual probability of this is

$$\Pr(N\ \text{data}) = e^{-\lambda \Sigma N \Delta h} \prod^{K} Y\,(N\Delta h) \qquad (3.2.3)$$

where Δh is the arbitrary sampling interval starting at $t = L/2$. The point of interest is that unless $\lambda = 0$ the probability of N data points is less than one and, hence, the probability of usable data is also less than one.

This is the worst case; if now we allow $l \rightarrow 0$, i.e., the lag to vanish or β to become large the number N of points required to interpret the data becomes less than in the marginal case above.

Similar considerations hold for metabolism experiments. In the early part of the experiment substrate is being converted to label in two or more culture chambers. At this point the different rates of label release may be due to irregularities in sample size and so forth. After some time has passed, however, inhibitors are added to one chamber and now it is the difference in the amounts of accumulated label from each chamber which is the quantity of interest. This quantity is likely to be most easily seen in the terminal phase of the experiment.

As a final example, in an organic analysis by gas chromatography it is the larger molecular weight species which are likely to be of the most biological interest. These species are eluted from the stationary phase at later times than the lighter fragments. This again leads to the requirement that the STR of the system must be held at a relatively high level during the later stages of the lifetime of the experiment.

3.3 Confounding and Aliasing

Confounding and aliasing are special sources of bias. The methods of elimination of these sources are as varied as the physical-biological principles of the various experiments. Thus, no general prescription for their detailed control can be offered here. One principle, however, does seem worthy of enunciation despite its rather obvious nature, and that is:

> In all experiments the bias control must be built into the experiment and explicitly monitored rather than attempting to use statistical legerdemain on the biased data after they have been collected. We cannot compute unbiasedness into data, we must build it in!

The degree to which an experimental implementation adheres to this principle is an important criterion for judging its potential worth.

4.0 COMPUTER SIMULATION ACTIVITIES

4.1 Introduction

Two simulations were programmed and run during the first year. These were designed to explicate some of the problems associated with obtaining a desirable degree of reliability in direct life detection experiments which observe growth or metabolism in a culture of Martian surface material. Although growth and metabolism experiments may be implemented in a number of ways, all of these experiments can be said to observe a change with time in the quantity or intensity of some property which is assumed either to index the size of a population of microorganisms in a sample of collected material (e.g., turbidity), or to indicate the occurrence of metabolism in such a population by observing the increase in concentration of some metabolic product. (Although increase in the size of such a population is not in principle necessary to the detection of such organisms, most of the experiments proposed seem to be designed to test the hypothesis that the population is actually increasing.)

If we assume that the property being observed actually does reflect changes in population size and/or the occurrence of metabolism, then the experiment can be said to test the hypothesis that changes in this observed property indicative of the occurrence of growth or metabolism do take place during the time period through which the observations are made.

The interpretation of the results of such an experiment requires a rule which permits the experimenter assign every observed sequence of measurements of the index property to one of two classes--that indicative of the presence of microorganisms and that not so indicative. The reliability of the resulting experiment is a function of the likelihood that false conclusions will be drawn, namely, that some observations will be interpreted as indicating the presence of microorganisms when, in fact, they are absent, or, conversely, that some results will be interpreted as indicative of the absence of microorganisms when, in fact, they are present.

Some of the many factors which may lead to false positive or false negative results are due to the kinds of bias and noise frequently encountered in analogous terrestrial experiments of this nature. Among these are:

1. Confounding due to the fact that fluctuations in the index property occur from abiological causes. For example, a radioactive metabolic product may be abiogenically released from the substrate provided, or the Geiger counter may record background radiation. Similarly, soil particles contribute to the measured turbidity.

2. Bias due to machine error may result from improper calibration or excessive drift to introduce an unanticipated systematic error. Other machine malfunctions can also introduce systematic errors.

3. Electronic noise in the measurement system introduces a random error and limits the precision of the measurements.

4. Biological noise consisting of random fluctuations in the population size due to the effects of uncontrolled variables may make the results appear erratic.

5. Transmission errors from noise in the communication system can corrupt the transmitted observations.

There are, in addition, certain other problems that may arise from the fact that the experiment as designed is not properly matched to the Martian microorganisms, so that even though live organisms are present in the sample cultured, their presence is not detected. Among these are the following:

1. The conditions of medium, temperature and so forth are wrong so that the organisms either fail to survive, or grow or metabolize at such a slow rate that they do not produce a detectable and significant change in the measured index property.
2. The organisms may, in fact, thrive and produce changes in the measured property which are normal for those organisms, but nevertheless are too slow or too rapid for the experiment to detect as a significant change. For example, metabolism may occur at a very slow rate, so that the increase in concentration of metabolic product during the experiment does not appear to be significant. (A similar result may be due to the fact that the innoculum was too small.) Conversely, the change may be so great relative to the length of the intervals between observations that the experimenter interprets the observations as bias due to machine malfunction.

The first computer simulation programed and run during the year simulated the growth and metabolism of a colony of Martian microorganisms. Entitled SOUP GEDANKEN, it is an imaginary experiment which is not closely modeled after any of the real experiments now under development. It was designed to explore in a limited way some of the interactions between some of the factors listed above which might be expected to affect all of the direct life detection experiments of the growth and metabolism type. The second is a simulation of a sampling system which was designed to explore some problems associated with obtaining inocula from samples of Martian surface material. It simulates a simplified surface matter collection system and permits an examination of effects of alternative assumptions regarding variations in amounts of material collected and in population densities per unit material on the resulting distributions of inoculum population sizes. These two programs are described separately below.

4.2 SOUP GEDANKEN

4.2.1 General Description

The experiment modeled in SOUP GEDANKEN is a culture experiment that is assumed to be performed by a Martian lander which detects the presence of microorganisms by measuring the concentration of a metabolic product released by a culture of such organisms grown in a suitable medium. Periodically the contents of this chamber are "observed" by a measuring device which measures the accumulated quantity of the metabolic product and then transmits the measurements to earth. The model was programed on a PDP-1 computer as a game played between a person acting the role of an experimenter and Nature as simulated in accordance with the model. The experimenter "designs" the experiment by specifying the range and sensitivity of the detector, the frequency of observations, the time at which observations are initiated and the total duration of the experiment. The experiment is then simulated and the experimenter is presented with the telemetered data points plotted on the face of the PDP output scope. He must then decide whether or not these results are indicative of the presence of microorganisms. He may change his design parameters and find out what results would have been transmitted had the new design parameters been in effect during the running of the experiment. He may also query the memory of the computer to find out what occurred during the course of the simulated experiment.

4.2.2 Components of the Model

1. Dimensions. The behavior of the elements modeled is expressed with reference to arbitrary time and quantity units. In particular, the time axis of the various plots is divided into 128 coordinate arbitrary time units, and the maximum duration of the experiment is 128 of these time units. They may usefully be thought of as corresponding to intervals of anywhere from one-half to four hours (depending on the shape of the idealized growth curve the population is assumed to approximate, see below). For convenience, the culture chamber may be thought of as having a volume of 100 cubic centimeters and the sample of Martian surface material introduced into it as having a volume of one cubic centimeter. The plots of population per unit time show the $\log_{10}$ of average population per cc. of culture medium, and the maximum value which can be plotted corresponds to an average concentration of 10^{14} organisms per cc.

2. Sample Collection and Growth of Population. This portion of the program generates a table of values of the population inside the culture chamber at each of the 128 experimental time units. What happens after the sample has been introduced into the chamber is determined with reference to a number of variables whose values may be assigned in advance by the user of the program. The procedures used to make these determinations and the relevant variables which govern the outcome are described sequentially below. However, one set of input values is used by many of the subroutines, and for this reason and because of its importance in determining the results of each simulated experiment, it will be described here. This is the idealized growth curve, which is a plot of log population per unit time describing the growth behavior which the Martian microorganisms would follow under idealized conditions of medium, temperature, etc. This plot can be manually entered by the experimenter prior to each run, merely by drawing it on the face of the PDP input-output scope. (Although this plot is entered and displayed on a 128-unit time axis, it merely shows the shape of the growth curve, as the length of time required by the population to cycle through the growth phases pictured may be as little as one quarter or as much as four times the 128 experimental time units, rate of growth being a random variable whose value is selected during the simulation run.) Three of the idealized growth curves which have been used are shown in Figure 1.*

*The decision to employ a hand drawn synthetic idealized growth curve rather than calculate the growth pattern with reference to an analytically expressed growth function was prompted by the fact that only a limited variety of shapes can be generated by changing the values of the variables in such an expression. Many of the resulting curves might be biologically quite unrealistic if the expression were complex enough to account for a reasonably variable lag period or for the effects of competition between two or more species. Although the present procedure is cumbersome, it simplifies the problems associated with examining the effects of "unusual" growth curves.

4.2.3 Description of Program

The procedures used to determine whether or not organisms are present and if so how they grow are described below.

1. Sample Collection. The first decision which must be made is whether or not live organisms capable of surviving in the chamber for any time at all are present in the sample collected. A random coin routine with probability p_1 of "yes" (assigned by the user) is used to make this decision.

If some responsive organisms are present, they may be too few in number, or too alien to grow well in the medium provided, and a degenerate colony will result. The probability that a degenerate colony results is p_2. If a degenerate colony occurs, it may die off entirely at any time t, with probability of dieing at time t being $(1-p_4)^t$ where the value of p_4 is assigned by the user. Until it dies off, the population of the degenerate colony will vary randomly between one and ten organisms per cc. of medium. An empiric growth curve which plots the simulated growth of the organism is generated and stored.

If viable organisms are introduced into the chamber and a degenerate colony does not result, then the program must determine how large the inoculum is and how the resulting colony grows. The maximum concentration of organisms in the one cc. sample is assumed to be one-half the maximum concentration shown in the idealized growth curve. The minimum concentration in the sample is assumed to be one quarter of this quantity or $10^{2.5}$, whichever is smaller. This range of sample concentrations is then divided into ten equally spaced increments which correspond to ten equally likely inoculum sizes. A random number generating routine is used to determine which of the ten sample sizes has occurred. This number varies between zero and nine, large values of the number corresponding to smaller inoculum sizes. The resulting population is divided by 100 in order to calculate the average population per cc. of medium present in the chamber immediately after introduction of the sample, and this is entered as the first value in the empirical growth curve.

The next decision which is made determines whether the culture grows rapidly or slowly. Six equally likely growth rates labeled zero through five are possible, and the effect of the selected growth rate is to "shrink" the time axis on which the idealized growth curve is displayed by a factor of four or two (in case the growth rate zero or one is chosen); to "stretch" it by a factor of two, four, or eight (in case the growth rate chosen is three, four, or five is chosen); or to leave it alone, which occurs when the factor two is chosen.*

*The effects of this wide range of possible growth rates may be interpreted as providing, e.g., for an average generation time as little as fifteen minutes or as long as eight hours. Thus, if each experimental time unit is interpreted as being equal to one hour, then the difference between the fastest and the slowest growth rates may be expressed by saying that the growth exhibited by the most rapidly growing colony in one hour requires thirty two hours for the most slowly growing colony to exhibit. (Whether any colony actually doubles its population in one experimental time unit depends on how the idealized curve was drawn.) The assumption that the minimum and maximum rates differ by a factor of thirty two may not be biologically unrealistic. However, the division of this range into only six rates, each of which is half the rate of the next slowest, and each of which is equally likely, is less realistic. For example, if the fastest growth rate is interpreted as providing for a minimum generation time of one hour, then the slowest is thirty two hours and the next slowest is sixteen, the next is eight, and so forth; there is no rate corresponding to ten hours, none to twenty or twenty-five. The advantages of simplicity resulting from this biologically unrealistic feature of the model may be defended on the grounds of the programing simplicities achieved and the fact that the model has not been designed to predict the likelihood of success of an experiment as a function of assumptions regarding a more realistic set of growth rates and a more realistic probability density distribution defined over them.

The program next calculates the duration of a lag period during which the population in the chamber remains fixed and equal to the population present in the sample collected. The duration of the lag depends on the size of the inoculum, being shorter for larger inocula, and on the chosen growth rate, being longer for slow growing species and correspondingly shorter for the faster growing species. The total lag period L is calculated as:

$$L = L_m + rk + gwk \quad \text{where}$$

L_m is a minimum lag period, r is the random number in the range zero to nine which determines the inoculum size, k is a lag factor set by the user, g is the growth rate factor in the range zero to five, and w is a weighting factor used to determine the relative effects of growth factor vs. inoculum size on the total lag period. All the runs reported in this document used one as the value of L_m and two as the value of k and w. This assignment yields a minimum lag period of one experimental time unit (for the largest inoculum and the fastest growth rate) and a maximum of forty three for the smallest inoculum and the slowest growth rate, which corresponds to approximately one third of the maximum duration of the experiment. During play, more than half the time the lag period is less than twenty time units.

Having calculated the lag period and the initial population density, the program now calculates a "modified idealized growth curve" consisting of the calculated lag period during which the value of the population is the same as the inoculum, and that portion of the stretched or shrunk idealized growth curve beginning with those values which are equal to or greater than the calculated initial density. (Thus, if the experimenter had included a lag period in his ideal curve showing slow growth when the population is small, this lag period would not be repeated if the initial population density was not that small.) Some modified growth curves showing the effects of calculated lags and changes in growth rates are shown in Figures 7 and 13. Figure 7 shows the modified growth curve resulting from combining an idealized growth curve providing for relatively rapid growth and high maximum population density (curve 1 in Figure 1) with a large inoculum and fast growth rate ($r = 0$, $g = 0$). Figure 13 shows a modified growth curve generated from the slowest growing idealized growth curve in Figure 1 with a small inoculum and slow growth rate ($r = 9$, $g = 4$). The modified growth curve contains 128 values corresponding to the 128 experimental time units. (If necessary, the final values of the idealized curve are extrapolated to the end of the time period.)

The empirical growth curve, which shows how the simulated colony actually grew, is calculated with reference to this modified growth curve. It departs from the latter mainly in showing the effects of a random error corresponding to biological noise. The first L values of the empirical growth curve are identical to the initial density, as these correspond to the lag period during which no change in population is observed. The remaining values all show the increment which occurs during the corresponding interval in the modified idealized curve, but they are also perturbed by a random error with mean zero and standard deviation equal to a constant H supplied by the user, and the frequencies of the resulting fluctuations are smoothed by means of a digital filter. Since population values (on all curves) are expressed as logarithms, this means that the rate of change will vary randomly from that exhibited in the idealized curve. The value of H used in most of the runs shown in the accompanying figures was .125 which corresponds to rate deviation of $\pm$ 33 per cent from that exhibited in the corresponding modified growth curve. The only restriction on this biological noise is that the resulting population density cannot exceed the maximum value shown in the (original) idealized growth curve. This restriction has the effect of forcing the population to decline after it has reached the maximum value even when no death phase is shown in the idealized growth curve.

2. Production and Measurement of Metabolic By-Product. This portion of the program simulates the behavior of a detector which measures the accumulated amount of metabolic product produced by the population of microorganisms. This is effected in two steps. The first step consists in determining how much of the product has accumulated at the beginning of each experimental time period. This is equal to a constant times the sum of the populations present during each of the preceeding time intervals. The response of the discriminator is assumed to be logarithmic, and in this model, the output at time t is proportional to:

$$f_t = \log_{10} \sum_{j=0}^{t} c \; 10^{E_j}$$

where E_j is the value of the empirical growth curve at time j and c is a proportionality constant which has been assigned the value 1.

Each detector has a certain sensitivity which determines the minimum change in the metabolic product concentration which can be detected. It also has a range which is determined by the minimum non-zero or threshold quantity which can be detected, and the maximum or saturation quantity which can be detected. The output of the detector varies within this range in increments determined by the sensitivity. For convenience in plotting the output of the detector and comparing it with the plots of the population, the maximum output value which can be shown is 10^{15}. (Since the maximum average population density is 10^{12}, the concentration of metabolic product never exceeds 10^{15}.)

The individual playing the role of experimenter is allowed to "design" the detector by assigning it a threshold value 0_{min} and a saturation value 0_{max}. He also specifies the sensitivity by assigning a value to i_s which corresponds to the number of intervals in the range 0_{min} to 0_{max} through which the output of the detector may range. When $i_s = 2^n$ exactly n binary digits are required to code the value and thus the number of bits required to transmit all the observations is equal to the product of n and the number of observations. The experimenter must specify the number of bits required for each measurement. In calculating the output of the discriminator at each time unit, the effects of a digitizing error are simulated by adding an error term to the calculated value of f_t. The digitizing error e_d is gaussian normal with mean 0 and standard deviation equal to z where:

$$z = [(0_{max} - 0_{min})/i_s]/\sqrt{12}$$

In order to simulate the digitizing of the output, the sum $f_t + e_d$ is rounded off to exactly n binary places. If f_t is less than 0_{min} (that is, if a detectable quantity has not yet accumulated) then 0_{min} is stored as the output value; correspondingly, if f_t is greater than 0_{max}, then 0_{max} is stored as the output value. It will be noted that the detector is "noise free" in the sense that no electronic noise is assumed to exist. Similarly, no confounding in the form either of noise or of a systematic error explicitly is modeled.

3. Data Transmission. This portion of the program simulates the performance of the data transmission system and the corruption of the transmitted results due to noise in the communication system. It does so by generating a table of 128 values of the telemetered messages which are the same as the 128 values of the detector output except that some bits are reversed. The probability that a bit may be reversed is independent of the bit position and is equal to p_t, which is assigned a value by the user. (It is usually convenient to suppose that assigning this value is one of the options of the experimenter. A value of .01, corresponding to the rather high reversal rate of one per hundred transmitted bits has been used in all the runs reported here.)

4. Data Output. The design options which determine the total number of observations, the duration of the experiment and the time at which the first observation is made are simulated by an output routine which selects the corresponding values of the message table for display. Thus, changing the "design" of the experiment in these respects does not require that a new culture be grown.

4.2.4 Representative Simulated Experiments

This program was designed as a game in order to demonstrate the effects of variations in growth rate, sample size, alternative ranges and sensitivities of the detectors, etc. The purpose was to focus primarily on problems represented by "ambiguous" results that might be difficult to interpret, and not to provide a mechanism for predicting the likelihood that such ambiguous results would occur. Figures 1 through 16 are plots of representative outputs of this program. The effects demonstrated in these plots are discussed below.

1. Biological Noise. Although the value of the variable which determines the effects of biological noise was relatively low for all simulated runs plotted, it resulted in fairly erratic empirical growth curves. For example, Figures 2 and 3 show the empirical growth curves generated in two runs where the idealized growth curve, the inoculum size and the growth rate were held constant, so that the results were as "identical" as the model permits. The value of H used in these runs was .25, corresponding to a mean deviation of ±78 per cent from the growth rate exhibited in the modified ideal growth curve. It will be noted that the variation due to noise is independent of the phase of growth, except that it does not occur during the lag period.

If this biological noise model is a reasonably realistic one, then it would seem that this noise source would seriously limit the value of control cultures. (More serious limitations on the use of such controls are likely to arise from the fact that the initial populations would be difficult to make equal. Problems of this nature were explored in more detail with the second simulation program described below.)

It will be noted that although the actual population fluctuated somewhat erratically, this fluctuation was not greatly apparent in the telemetered results; and the telemetered observations obtained from these two experiments look very much alike. This arises from the fact that the index property is the accumulated quantity of metabolic product: summing this product damps the effects of biological noise. If the index property were the rate of production rather than the sum produced, the detector output would reflect these fluctuations. The remaining empiric curves plotted in the illustration were calculated with an H value of .125, corresponding to a mean deviation of ± 33 per cent.

2. Confounding. No confounding factors were explicitly modeled in this program, in that confounding effects were not added to the outputs either when no viable organisms were introduced or when growth of a non-degenerate colony was simulated. However, confounding due to random abiological liberation of the labeled metabolic product in small quantities would have an effect identical to that of a degenerate colony which survived for the duration of the experiment. Figure 4 shows the empiric growth curve for such a colony together with telemetry points showing the gradual accumulation of the product as it would be detected by a sensitive detector with a wide range (.5 log unit sensitivity and $O_{min} = 10^{.5}$). Figure 5 shows the telemetered data which would result if the detector were less sensitive and possessed a higher detector threshold (1 log unit sensitivity and $O_{min} = 10^{2}$).

3. Effects of Sample Size, Growth Rate and Detector Function. The combined effects of large sample sizes and rapid growth rates serve to saturate the detector rapidly, by driving the concentration to its maximum detectable value. How rapidly this occurs also depends upon the ideal growth curve employed and the value of 0_{max}. Figures 7 through 12 show the effects of a large sample and fast growth combined with idealized growth curve 1, which has a large maximum value. Figure 7 shows the modified idealized curve and the empirical growth curve. This empiric curve was used in all results plotted in Figures 8 through 12. Figures 8, 9, and 10 illustrate what would be received with a wide range, high sensitivity detector but with different observation frequencies. It will be noted that Figure 10 also corresponds to what would happen if a machine error resulted in rapid and excessive drift of the detector output. This example shows the value of providing frequent observations and independent monitoring of the detector so as to aid in diagnosing machine errors of this kind.

The effects of alternate sensitivities and ranges are shown in Figures 11 and 12. Figure 11 shows the data points resulting from medium sensitivity detector (1 log unit) with wide range. Figure 12 shows the data which would be received if the detector had a low saturation value (10^{11}).* Figures 11 and 12 look even more like excessive drift than Figure 10.

*None of the simulated experiments plotted here involved a maximum population density greater than 10^{11}, and consequently the maximum concentration did not exceed 10^{12}. If an idealized growth curve with higher values had been employed, the misleading effects of saturation would be more marked.

If the inoculum size is small and/or the growth rate is slow, growth would be very difficult to distinguish from confounding which might be attributed to abiological sources. Idealized growth curve 3 in Figure 1 shows a very gradual growth such as might be characteristic of organisms not well adapted to the medium provided. This curve is combined with a small inoculum and a moderately slow generation time in Figure 13, which shows the resulting modified and empiric growth curves. Figure 14 shows the telemetered points which would result from this experiment if a sensitive detector with a low threshold were employed. Figure 15 shows the same experiment as it would be observed with a detector of medium sensitivity (1 log unit) and higher threshold (10^2).

Figures 14 and 15 strongly resemble the data points plotted in Figures 4 and 5, which suggest slow growth would be very difficult to distinguish from confounding alone if the experimenter could not anticipate the level of confounding which would occur. It is obvious that in this slow growth, small inoculum case, shorter durations would fail to detect anything at all.

4. Detector Sensitivity and Range. No detector noise or drift is simulated in SOUP GEDANKEN, but the sensitivity and range combinations modeled may usefully be interpreted as choices forced on the designer of a detector system as alternatives to detectors which combine the advantages of greater sensitivity-range features with the hazards of increased noise and liability to drift frequently associated with such systems. It will be noted that reducing the sensitivity of the system modeled here (where the output varies linearly with the log of the index

property, which, in turn, varies with the integral of the population density), makes it quite insensitive to changes in the population densities which occur during the end of the experiment when the concentration has reached a level near the high end of the scale. For example, in Figures 8 through 10, the concentration does not change after the 48th time interval even though the population continues at a high density. The output would remain at this level even if the whole population were to suddenly succumb. Narrowing the range of a detector system sometimes makes it possible to increase the sensitivity, but this has certain disadvantages such as those pictured in Figure 16, where the output shows either the minimum or maximum value during almost 7/8ths of the duration of the experiment. In effect, the narrower the range, the fewer the combinations of inoculum size and growth rate which can be detected.

5. Communication Noise. The probability of bit reversal modeled in all of these simulated experiments is relatively high. In spite of this fact, and in spite of the fact that the probability is independent of bit position (so that relatively disastrous reversal of the most significant bits are just as likely as less consequential reversals of insignificant bits), communication noise does not greatly affect the interpretability of the results. For example, most of the accompanying plots show bit reversals in significant positions, but they are easy to spot as such. From this it may be concluded that any additional communication capacity that may be available would be more effectively used in providing for more frequent observations or in monitoring of machine function than in

repeated transmission of the same observational data. (That is, the redundancy implicit in greater observation frequency is seen in these examples to be more efficacious than that effected by error-correction coding or duplicate transmission.)

4.2.5 Evaluation of Results

SOUP GEDANKEN models a hypothetical experiment which is extremely simplified in concept and function. As a demonstration model of some possible effects of biological noise and mismatch between the experiment as designed and Martian reality as it might be encountered, it appears to have some value, much of which arises from the fact that it can be played as a game which observers find interesting and convincing. However, it has very limited value as a prediction model, because the validity of the assumptions has not been determined, and because a very large number of serious possible sources of ambiguity are not represented in the model. There is, for example, no machine malfunction which gives rise to bias, no electronic noise, and no confounding. The range of inoculum sizes played is, in fact, very small, and the range of growth rates might be considered excessive due to the fact that the duration of the experiment is arbitrarily limited to 128 units while the slowest growing population could require as much as 565 time units to complete the growth phases pictured in the idealized growth curve. (It should be pointed out, however, that to assume that all growth will necessarily take place during the duration of the experiment is tantamount to assuming that the experimenter is always either perfectly lucky or perfectly informed about Martian conditions.) No attempt was made to make SOUP

GEDANKEN more realistic or more sophisticated than it is, because it is not possible at this time to verify the validity of the assumptions which would have to be incorporated in a more "realistic" model.

Nevertheless, SOUP GEDANKEN does provide valuable qualitative and semiquantitative insight into the non-linear behavior of interaction between the shape of the growth curve, the growth rate, inoculum size and biological noise. Some valid conclusions can be drawn from these simulated experiments. Among these is the fact that the shape of the growth curve cannot readily be inferred from observations of the quantity of accumulated metabolic product. This being the case, many types of confounding will be difficult to discriminate from the effects of a population of organisms. Designers of life detection experiments have recognized this problem, the obvious solution to which is the employment of controls by simultaneously conducting many experiments which differ in controlled ways, and in particular, in using inhibitors to kill the population in some cultures and to compare the results with cultures which were not treated with inhibitors.

4.3 SAMPLER

4.3.1 General Description

The second simulation program designed and run during the course of the reporting period simulates the behavior of a pair of sampling systems which function in parallel to collect two samples of Martian surface material which will then be introduced into two culture chambers. Although the program was written so that many of its components could be reasonably interpreted as representing certain machine functions performed by an early model of the Gulliver sticky string collection system, the simulation model is a relativey general one and can, under certain assignments of values to its variables, be interpreted as representing a variety of other possible collection systems. The model is a stochastic model in that all of its inputs are probability density distributions which govern the likelihoods of occurrences of the various events simulated, and the output consists of a statistical analysis of the events observed during a large number of runs at each set of input values.

The model was designed to explore two problems associated with growth and metabolism experiments. The first of these is that represented by the effects of variability of innoculum size on the reliability of such experiments. As was pointed out in the preceding discussion of SOUP GEDANKEN, the larger the innoculum, the more likely it is that population growth will occur and be detected during any fixed duration of the experiment. The relationship between innoculum size and likelihood of increases in the population is stronger (and more critical) the more the culture conditions differ from the normal habitat of the responsive organisms. All growth and metabolism experiments involve some tampering with the local environment (even the "bell jar" experiments which do not involve a medium of terrestrial origin) and the percentage of organisms which will survive the tampering and adapt to the new conditions may be reasonably assumed to be fixed. Therefore, the smaller the population in the sample, the smaller the number of adaptable organisms in the innoculum and hence the longer it will take for the adapting population to reach a detectable threshold.

Because the range and sensitivity of the detector and the duration of the experiment may be reasonably supposed to be somewhat limited, it is of some interest to know how the innoculum population may be expected to vary under various assumptions regarding the mean and variance of the density of responsive organisms per unit to collected material, and related assumptions regarding the variability of the quantities of material collected.

The second problem which the model was designed to explicate concerns the use of controls involving diversity of treatments (especially inhibition of growth) applied to cultures which are otherwise as similar as possible. If the populations of organisms introduced into two otherwise similar culture habitats differ greatly in relative magnitude, then one of these resulting cultures cannot be used as an effective control for the other. Consequently, in evaluating the utility of any particular control scheme, one of the important factors to consider is the probability that the two populations will differ in magnitude by some predetermined amount which may be interpreted as corresponding to the maximum permissable variation consistent with "good control".

Several sample collection procedures have been proposed for inclusion in Martian landers. Some of these involve the deployment of a single collector system which gathers a quantity of material and then introduces aliquots of this sample into each of a variety of culture chambers. Others involve the deployment of two or more relatively independent collection systems, each of which gathers material, where all or some measured fraction of the material gathered by each system is used to the innoculum for a separate culture. In all cases the actual population of responsive organisms introduced into each chamber cannot be completely controlled so as to be held constant, since the amount of collected material introduced into each culture chamber (or otherwise subjected to observation) can always be expected to differ by some amount no matter how carefully it is measured, and furthermore it will not be possible to insure that the population density per unit of material will be equal even if quantities of material are reasonably similar. Some variation is therefore inevitable, and the sampler program described here was designed to examine some effects of the interaction of a number of factors that can reasonably be said to correspond to identifiable sources of variation.

4.3.2 Components of the Model

The components of the model will be described here in terms which show their relationships to the sticky string system which the model can be said to simulate. Alternative interpretations in terms of other kinds of systems will be presented in the discussion of results.

The model may usefully be conceived of as consisting of two major components, the first of which determines the manner in which the machine functions are sequenced and executed, while the second determines the effects which may be attributed to "Martian" variables. The behavior of the machine portion of the program consists of a sequence of control functions assumed to be performed by the programming unit of the sampler system, intermixed in time with a sequence of mechanical functions executed by the mechanical portions of the unit. Mechanically, the system may be thought of as consisting of two string units and of their associated units which perform the necessary collection functions. These functions effect deployment of the string by firing a projectile which causes it to be unwound and strung out across the surface of the ground. Each string has an independent firing system. After firing, a common windup motor is started, causing both the strings to be wound back on their spindles, gathering surface material as they are dragged over the ground. Then an ampule-breaking mechanism associated with each string system is activated in order to break an ampule located inside each spindle containing a prepared medium. Breaking the ampule starts the culture experiment by "introducing" each innoculum collected by the string into the liberated medium.

The mechanical functions associated with deployment, windup, and ampule breakage are initiated in their proper sequence by a control unit which emits appropriate signals. The computer program begins with an attempt by the sequencer to emit a signal causing the projectile of the first string to fire. The probability that this signal will be emitted is P_{s_i} (the initial sequencer functioning probability) which is one of the input values of the model. If the sequencer fails to function, P_{si} is decremented by a fixed amount m so that the probability of subsequent sequencer function is correspondingly reduced. The size of m determines in part the degree of integration of the whole system. All five of the mechanical functions described below require a sequencer signal for their initiation. If a sequencer signal is not available for a specific function, that function is skipped and no further operation associated with that string is attempted. (If the windup motor fails, the experiment is terminated.) An attempt is made to execute the next operation associated with the other string system. The total system is said to fail completely only if neither string unit completes all of its programmed functions (either through unavailability of the requisite sequence signals, or mechanical failure on the part of mechanical units).

The five mechanical functions programmed, and the associated variables which determine their probability of correct functioning are:

1. Firing of projectile associated with string 1; P 1
2. Firing of projectile associated with string 2; P 2
3. Windup of both strings simultaneously; Pw
4. Breakage of ampule 1; Pa_1
5. Breakage of ampule 2; Pa_2

If all the machine functions associated with one or both of the two string units are appropriately performed, the size of the innoculum will depend on the following:

1. Whether or not responsive organisms exist in the surface material in the vicinity of the lander. The probability of occurrence of responsive organisms is P (RMO).
2. The length of string deployed. The string may break if the projectile shoots out with too much force or if it gets snagged on a protruberance on the surface. Consequently, it is divided into a number of segments separated by weak connections so that if strain is placed on the string it will part at a weak place rather than become completely torn off or hung up. The string may be 0. 1, 2, 3, 4, 5, 6, or 7 units in length. The relative

probabilities of each of the resulting string lengths is determined with reference to a table of string length distribution and hence string length variation may be controlled by appropriate assignment of values to this table.

3. The adhesion factor associated with surface material in the vicinity of the lander. This factor may take any value from 1 to 10, and each value is equally probable. The total number of unit quantities of material collected by each string is determined by equation 1.

$$(1) \quad \text{xgrams collected} = a\ L_n^2$$

where a is the adhesion factor, L_1 is the length of string reeled in by the first string system and L_2 is that reeled in by the second system. The quantity is expressed in "xgrams" which may be interpreted as milligrams, centigrams, decigrams, etc. depending on how one chooses to interpret the significance of the density factor described below.

4. The number of responsive organisms per unit material collected is 10^d, where d, the density factor, varies normally about a mean delta with standard deviation lambda (where delta and lambda are variables which may be assigned values in the range 0 to 10). These two variables are employed together with an appropriate random number generating routine in order to select a different value of d for each string at each run. The population collected is thus $aL^2\ 10^d$.

4.3.3 Interpretation of the Model and Analysis of Results

This program was completed late in the reporting period and consequently the model it represents has not been thoroughly studied. The problems which it has been used to study and the results obtained to date are described below.

4.3.3.1 Variations in Innoculum Size

SOUP GEDANKEN illustrates one of the more serious types of miss-match which might occur between the experiment as designed and the Martian conditions encountered. This consists of a detector range and sensitivity which is inappropriate to the innoculum size-growth rate combination found. Consequently SAMPLER was employed to find out what ranges of population sizes might be encountered under various assumptions about the distribution of population densities as determined by assigned values of delta and lambda, and the variations in amounts of material collected as determined by the distribution of string lengths.

SAMPLER was used to simulate one thousand collections at each of 9 pairs of values of delta and lambda for each of five string distributions shown in Table 1. Because of the machine function probabilities assigned (see below) 98% of these runs resulted in "successful" collections. Each table corresponds to a fixed string length distribution, and the interpretation of the significance of these simulated collection runs relative to a given real collection system depends upon the extent to which the assigned values of the variables yield reasonably realistic representations of the behavior of the actual system.

The model provides that the quantity of surface matter gathered varies with the square of the string length, and this is believed to be a fairly realistic representation of the behavior of a real sticky string system. The collection behavior of such a system may be viewed as similar to that which would result if a large number of collection cells were distributed along the length of the string. The amount of material collected by each cell increases with the distance over which it is dragged while being reeled in. Thus, longer string lengths correspond to larger numbers of cells, and to longer drag distances for the more distant cells. According to this view of the collection behavior of a string, the amount collected will be an exponential function of length, the value of the exponent depending upon the maximum cell capacity and the average drag distance required to fill each cell to capacity. Although few experimental data are available, it appears that the range in amounts actually collected by real string systems actually exceeds those resulting from the length distributions employed here. Distributions 3 and 5 are believed to be the best representations of sticky string behavior, since they are skewed in a manner which would be the case if the string were rarely shortened by more than one or two units. They differ in respect to maximum length, and distribution 5 provides for 0 length with probability .01; this results in a difference in the total range and in the mean and standard deviation of xgrams collected (See Table 4.1).

The validity of the use of the non-linear factor L^2 has not been established for vacuum cleaner or claw collection systems. It is believed that none of the sample collection systems under development gather material in a manner which can be described as a linear function of an adhesion factor. Distributions 1, 2, and 4 can be said to represent varying degrees of control that might be exercised over the quantities collected and/or the division of a single sample into a number of aliquots. The results of simulated collection runs at these values are provided primarily for purposes of comparison with the more realistic string interpretation of the model.

Tables 4.2 and 4.3 show the results of varying delta and lambda in terms of the resulting means and 50% confidence limits on the number of organisms collected. Table 4.2 shows the actual population means and ranges with all values scaled to 10^5 in order to simplify comparisons. Table 4.3 shows the confidence limits represented as a multiple of the mean.

If we interpret xgrams as corresponding to milligrams, then it will be seen that delta = 6 corresponds to relatively heavily populated garden soil, delta = 4 represents relatively thinly populated areas and delta = 1 represents an unusually barren surface. Lambda values specify the homogeniety of population in the vicinity of the lander. Although population densities may be supposed to vary along the path traversed by the string much more than would be indicated by any of the values of lambda employed, the mixing effect of dragging the string across the ground so that it samples populations all along its path could be supposed to reduce population variance within the collected sample considerably. Lambda = 0.5 is therefore not an unreasonably low value, although lambda = 1.0 may be encountered in many areas on earth. The smallest value of lambda simulated is 0.1, and it is believed to represent an unrealistic degree of homogeniety, such as could be achieved only if a single sample were carefully mixed and then subdivided into two aliquots for two different cultures. An examination of the results shown in Tables 4.2 and 4.3 reveals the following:

1. Although the mean population values at each pair of delta-lambda values vary with maximum string length (as would be expected), the ranges of values in the 50% confidence inteval are insensitive to length distributions.

They average about 4 to 1 for lambda = 0.1, 8 to 1 for lambda = 0.5 and are greater than 25 to 1 for lambda = 1.0.

2. There is relatively little overlap among the ranges of population values resulting from different assignments of delta.

3. The population means and the ranges in the 50% confidence interval do not vary greatly with string length distribution if delta and lambda are held constant. The means do not vary by more than about 2.5 to 1.

4. The means and ranges vary greatly as a function of delta. Since this is true even for string length distribution 1 (corresponding to fairly close control over the quantity of material collected), one may conclude that growth and metabolism experiments ought to be designed so as to accommodate great variations in innoculum population size.

4.3.3.2 Population Variation and Effectiveness of Controls

Although the effects of innoculum size on duration of the lag period are not, to our knowledge, well understood, it can be expected that the time required for a "responsive" culture to reveal its presence varies with the size of the innoculum. Effective control is achieved only when the innoculum populations do not differ greatly in magnitude. Therefore, one figure of merit that may be applied to collection systems is the amount by which the populations in two simultaneously collected samples can be expected to differ. Table 4.4 shows the ratios of the larger to the smaller population observed in more than 480 simulated two-string collection experiments run at each pair of delta and lambda values. The smaller of the two tabulated figures is the estimated 50% confidence limit: fifty per cent of the pairs of populations collected will differ a factor greater than this value. The larger of the two tabulated values corresponds to the 68% confidence limit: 16% of the collected pairs of samples will differ in magnitude by a factor at least this great.

If we assume that control is "good" if the populations differ by not more than a factor of 4, "marginal" if the populations differ by not more than a factor of 12, and otherwise "inadequate", then examination of Tables 4.4c and 4.4c reveals that good or marginal control is achieved in more than 50% of the simulated runs with the modeled sticky string system. However, for any given set of values of delta and lambda, the probability of inadequate control is at least .16 (as revealed by the tabulated 68 per cent confidence limit).

It will be noted that although the effects of string length distribution are not great, distributions 3 and 5 appear to exhibit somewhat greater control effectiveness than the other distributions. Distribution 3 (corresponding to the longer sticky string system) shows greater control effectiveness than distribution 5 (which has a shorter maximum string length) primarily because the mean population is larger for the former and the latter includes some strings of 0 length.

4.3.3.3 Effects of Machine Variables

The machine portions of the simulated experiment are the sequencer and the five mechanical operations which must be performed if an attempt to gather two samples on the two strings is to be successful. The model provides for two kinds of "integration" of machine functions. The fact that one mechanical unit--the wind-up motor--is common to both string systems provides a limited degree of integration in that failure of this unit to operate results in failure of both string systems. The remaining two mechanical operations associated with string collection--the firing of the projectiles and the breaking of the ampules--are independent in that machine malfunction in one of these units does not affect the probability of functioning of the corresponding unit of the other system. If the program control unit operates perfectly to provide a sequencing signal every time one is needed, then the effects of assigned probabilities of operation of each of the five mechanical components may be examined. Table 4.5 shows such effects when each of the mechanical components is assigned the same probability of functioning. Five hundred simulated runs were conducted at each set of values. The frequency with which at least one string was successfully deployed is shown in column 1; the frequency with which both strings were deployed is shown in column 2. The effects of having both string systems share a common mechanical unit are exhibited in columns 3 and 4. Column 3 shows the frequency with which at least one string system successfully completed all operations; column 4 shows the frequency with which both string systems successfully completed all operations, so that a pair of cultures was started. The frequency with which the total system failed on account of failure to rewind the strings is shown in column 5, and the frequency with which the total system failed for any other reason is shown in column 6. The probability of successful operation of any single machine function may be supposed, in real systems, to be at least .99. However, four smaller values were also run because each of the five simulated functions represented in this model may be supposed to correspond to the operations of a number of distinct machine components whose joint probability of successful operation is likely to be less than .99. It will be noted that the frequency with which both collection systems operated successfully dropped off sharply with decreases in the probability of functioning of the five mechanical units modeled, in a manner characteristic of integrated systems.

The model of the sequencer employed provides control over another kind of integrating element. Every time the sequence unit fails to emit an appropriate signal, its probability of correct functioning is decremented by an amount equal to m. Large values of m correspond to control units with little redundancy in which there is a relatively high degree of correlation among failures. If the unit is sufficiently damaged to fail once, then it is likely to fail again in the near future. Small values of m correspond to relatively redundant systems which contain back-up control elements. Table 4.6 shows the effects of three different values of m when probability of machine operation and initial probability of sequencer function are set to .98. Columns 1 through 4 are the same as the corresponding columns of Table 4.5. Column 5 shows the frequency with which the wind-up motor failed to operate due either to mechanical failure or to failure to obtain a sequence signal. Column 6 shows the frequency at which the system failed because the sequencing unit had ceased to operate at all (probability of sequence function equal to or less than zero), column 7 shows the frequency of system failure due to all other causes and column 8 shows the mean value of the probability of sequencer operation at the end of the simulated experiment. It will be noted that the frequency with which both string systems successfully completed all operations does not vary greatly with m. It is obvious that if having two cultures is deemed essential to the success of the experiment, then redundancy of sequence control is not necessary; such redundancy is useful only when the experiment is viewed as at least partially successful when one string system functions successfully.

	PROBABILITY OF OCCURRENCE				
DISTRIBUTION	1	2	3	4	5
String Length					
0				.12	.01
1		.01	.01	.12	.05
2		.04	.02	.13	.09
3	.1	.25	.03	.13	.35
4	.4	.40	.09	.13	.50
5	.4	.25	.35	.13	
6	.1	.04	.50	.12	
7		.01		.12	
length	4.5 ± 0.8	4.0 ± 1.0	5.25 ± 1.0	3.5 ± 2.2	3.28 ± 0.9
xgrams	119 ± 73	96 ± 70	160 ± 102	130 ± 130	66 ± 46

TABLE 4.1

Five string length distributions and resulting mean and standard deviations of string length and xgrams collected.

Geometric Mean Biological Sample Load and 50% Confidence Limit.
All entries should be multiplied by 1×10^5

Table 4.2a Length distribution 1

δ \ λ	0.1	0.5	1.0
1	0.0091 0.0052 - 0.0158	0.0093 0.0034 - 0.0255	0.0096 0.0020 - 0.0471
4	8.95 3.99 - 20.1	9.47 3.58 - 25.0	7.74 1.55 - 38.5
6	852 477 - 1520	957 234 - 3910	920 176 - 4820

Table 4.2b Length distribution 2

δ	0.1	0.5	1.0
1	0.0066 0.0035 - 0.0124	0.0073 0.0026 - 0.0203	0.0064 0.0018 - 0.0346
4	6.47 3.37 - 12.5	6.91 2.51 - 19.0	7.90 1.52 - 41.1
6	682 367 - 1270	763 291 - 1990	676 130 - 3510

Table 4.2c Length distribution 3

δ \ λ	0.1	0.5	1.0
1	0.0119 0.0064 - 0.022	0.0116 0.0023 - 0.031	0.0137 0.0026 - 0.072
4	11.5 6.2 - 21.4	12.0 4.5 - 32.0	13.8 2.6 - 71.2
6	1160 647 - 2090	1140 420 - 3070	1260 238 - 6610

Table 4.2d Length distribution 4

δ	0.1	0.5	1.0
1	0.0066 0.0025 - 0.0172	0.0069 0.0021 - 0.0227	0.0059 0.0008 - 0.0400
4	5.63 2.08 - 15.2	6.30 1.82 - 21.9	5.76 1.05 - 32.1
6	600 219 - 1660	611 196 - 2670	603 104 - 3830

Table 4.2e Length distribution 5

δ \ λ	0.1	0.5	1.0
1	0.0044 0.0023 - 0.0086	0.0048 0.0017 - 0.0131	0.0048 0.0009 - 0.0265
4	4.26 2.25 - 8.08	4.58 1.75 - 12.0	5.02 0.95 - 43.9
6	458 231 - 906	457 179 - 1230	450 87.1 - 2109

Table 4.3 Fifty Per Cent Interval as Multiple of Geometric Mean Load

Table 4.3a Length distribution 1

δ \ λ	0.1	0.5	1.0
1	.58 - 1.73	0.36 - 2.73	0.20 - 4.91
4	.44 - 2.24	0.38 - 2.64	0.20 - 4.97
6	.56 - 1.78	0.24 - 4.08	0.19 - 5.24

Table 4.3b Length distribution 2

δ \ λ	0.1	0.5	1.0
1	0.54 - 1.88	0.36 - 2.80	.18 - 5.41
4	0.52 - 1.93	0.36 - 2.75	.19 - 5.20
6	0.54 - 1.86	0.38 - 2.61	0.19 - 5.19

Table 4.3c Length distribution 3

δ \ λ	0.1	0.5	1.0
1	0.54 - 1.85	0.20 - 2.67	0.19 - 5.26
4	0.54 - 1.86	0.38 - 2.67	0.19 - 5.16
6	0.56 - 1.80	0.37 - 2.69	0.19 - 5.25

Table 4.3d Length distribution 4

δ \ λ	0.1	0.5	1.0
1	0.38 - 2.60	0.30 - 3.28	0.14 - 6.78
4	0.37 - 2.69	0.29 - 3.48	0.18 - 5.57
6	0.36 - 2.77	0.32 - 4.37	0.17 - 6.35

Table 4.3e Length distribution 5

	0.1	0.5	1.0
1	0.52 - 1.95	0.35 - 2.73	0.19 - 5.48
4	0.52 - 1.90	0.38 - 2.62	0.19 - 8.74
6	0.50 - 1.98	0.39 - 2.69	0.19 - 4.69

Table 4.4 Ratios of larger to smaller populations simultaneously collected, 50% and 68% confidence intervals.

Table 4.4a Length distribution 1

	0.1		0.5		1.0	
	50%	68%	50%	68%	50%	68%
1	2.1	10.2	4.2	71.2	9.6	776.9
4	2.2	10.0	4.0	59.7	9.7	858.6
6	2.3	11.5	3.9	55.0	10.4	1024.0

Table 4.4b Length distribution 2

	0.1		0.5		1.0	
	50%	68%	50%	68%	50%	68%
1	2.6	14.8	4.1	71.3	11.2	1258.8
4	2.4	14.7	4.1	69.4	9.9	931.7
6	2.4	13.8	3.7	53.8	10.4	1020.4

Table 4.4c Length distribution 3

	0.1		0.5		1.0	
	50%	68%	50%	68%	50%	68%
1	2.3	12.7	4.0	61.6	11.2	1115.9
4	2.4	13.7	3.9	60.8	10.9	1013.4
6	2.3	11.0	3.8	57.2	10.2	1030.7

Table 4.4d Length distribution 3

	0.1		0.5		1.0	
	50%	68%	50%	68%	50%	68%
1	3.8	54.3	4.7	129.2	14.2	2757.9
4	4.1	61.8	5.4	176.0	10.1	1229.0
6	4.2	69.9	5.8	207.2	12.3	1904.0

Table 4.4e Length distribution 5

	0.1		0.5		1.0	
	50%	68%	50%	68%	50%	68%
1	2.3	14.6	4.6	78.2	10.9	1198.8
4	2.7	16.5	4.1	61.4	9.5	890.5
6	2.6	18.0	4.0	62.4	10.5	969.0

Probability of Operation	1	2	3	4	5	6
.99	.98	.962	.971	.936	.006	---
.98	.98	.942	.96	.899	.022	---
.97	.974	.918	.948	.856	.038	---
.93	.929	.811	.866	.696	.048	.008
.90	.903	.738	.816	.602	.102	.014

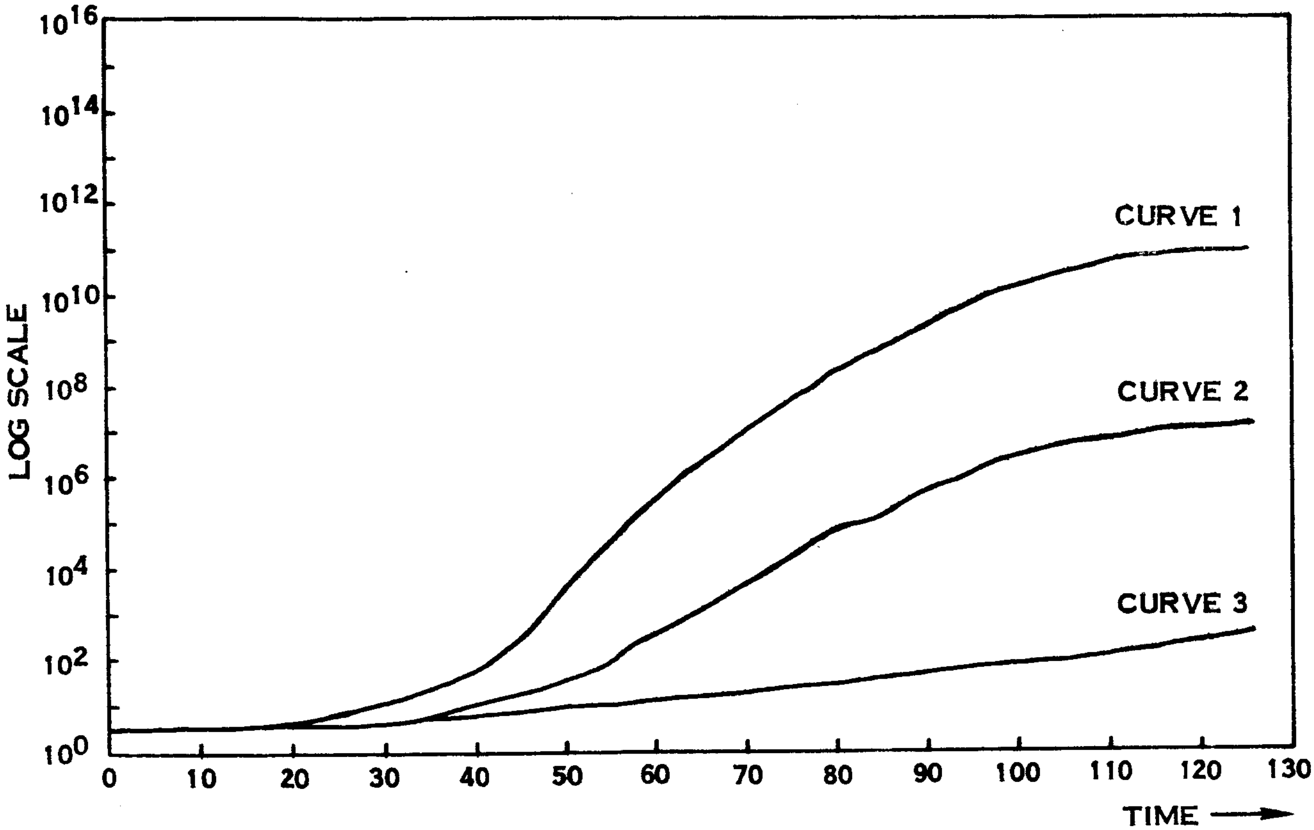
Table 4.5 Effects of Component Reliability on Total System Function. The sequencing unit was assumed to function perfectly. Tabulated probabilities of component operation were assigned to all five mechanical units simulated. Columns 1 through 6 show frequencies per thousand simulated experiments at which the following occurs:
Column 1, at least one string successfully deployed;
Column 2, both strings successfully deployed;
Column 3, at least one string system successfully completed all operations;
Column 4, both string systems successfully completed all operations;
Column 5, total system failure due to failure of wind-up motor;
Column 6, total system failure due to all other causes combined.

Value of m	1	2	3	4	5	6	7	8
.05	.966	.872	.932	.804	.052	----	.002	.974
.10	.952	.87[illegible]	.908	.792	.038	----	.004	.969
.50	.944	.847	.906	.[illegible]88	.042	.052	.002	.956

Table 4.6 Effects of Different Amounts of Redundancy in Control Unit Functioning.

Initial probability of sequencer function = .98, all mechanical function probabilities = .98, m is value by which sequencer function probability reduced after each failure, columns 1 though 7 are frequencies per thousand attempts at which the following occurred:

Column 1, at least one string deployed;
Column 2, both strings deployed;
Column 3, at least one string system successful;
Column 4, both string systems successful;
Column 5, wind-up motor failed to function (due either to machine failure or sequencer failure);
Column 6, sequencing unit ceased to operate;
Column 7, system failure due to all other causes combined;
Column 8, mean probability of sequencer function at end of experiment.

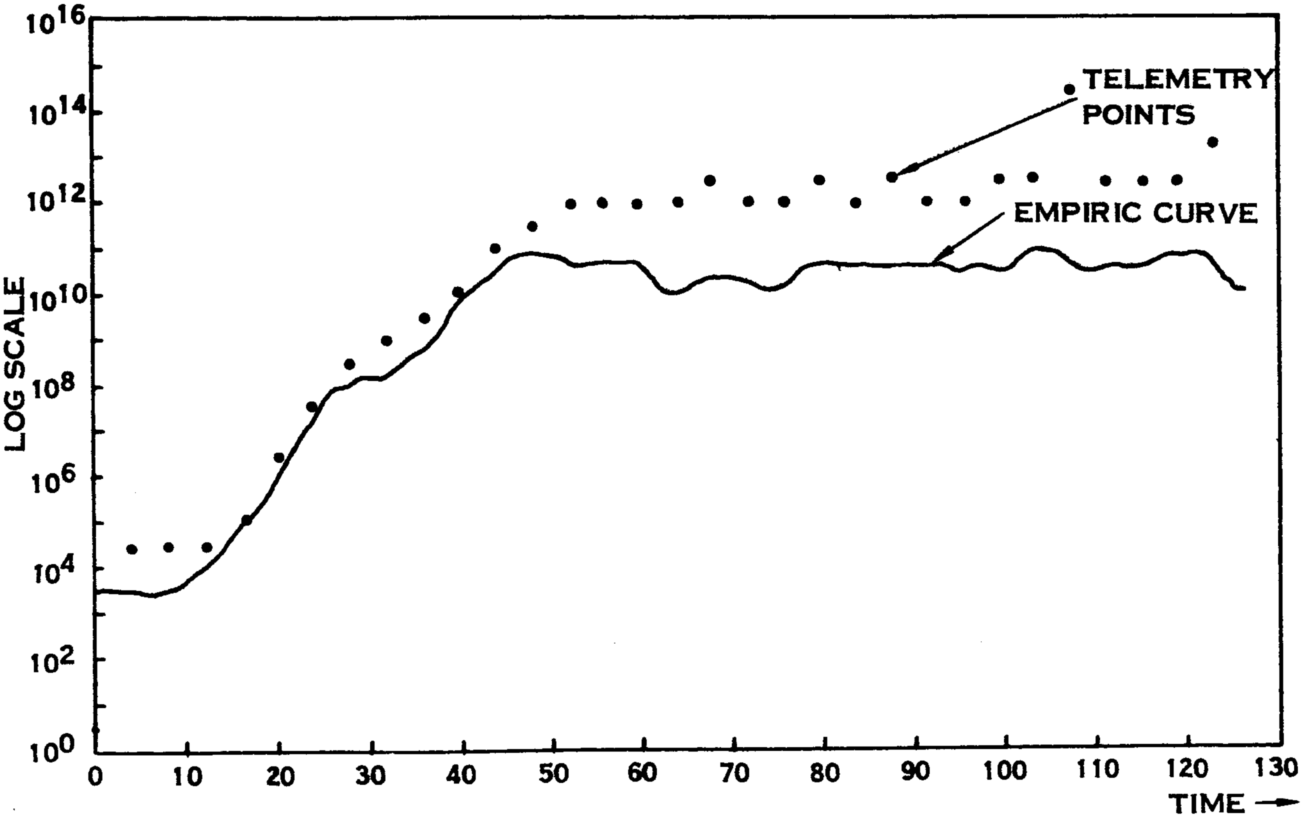


IDEALIZED GROWTH CURVES 1, 2 AND 3
FIGURE 1

INPUTS
IDEAL CURVE 1
$r = 0$
$g = 1$
$H = .25$

DETECTOR
SENSITIVITY = .5 LOG UNIT
RANGE 10^{-5} TO 10^{14}

32 TELEMENTED POINTS



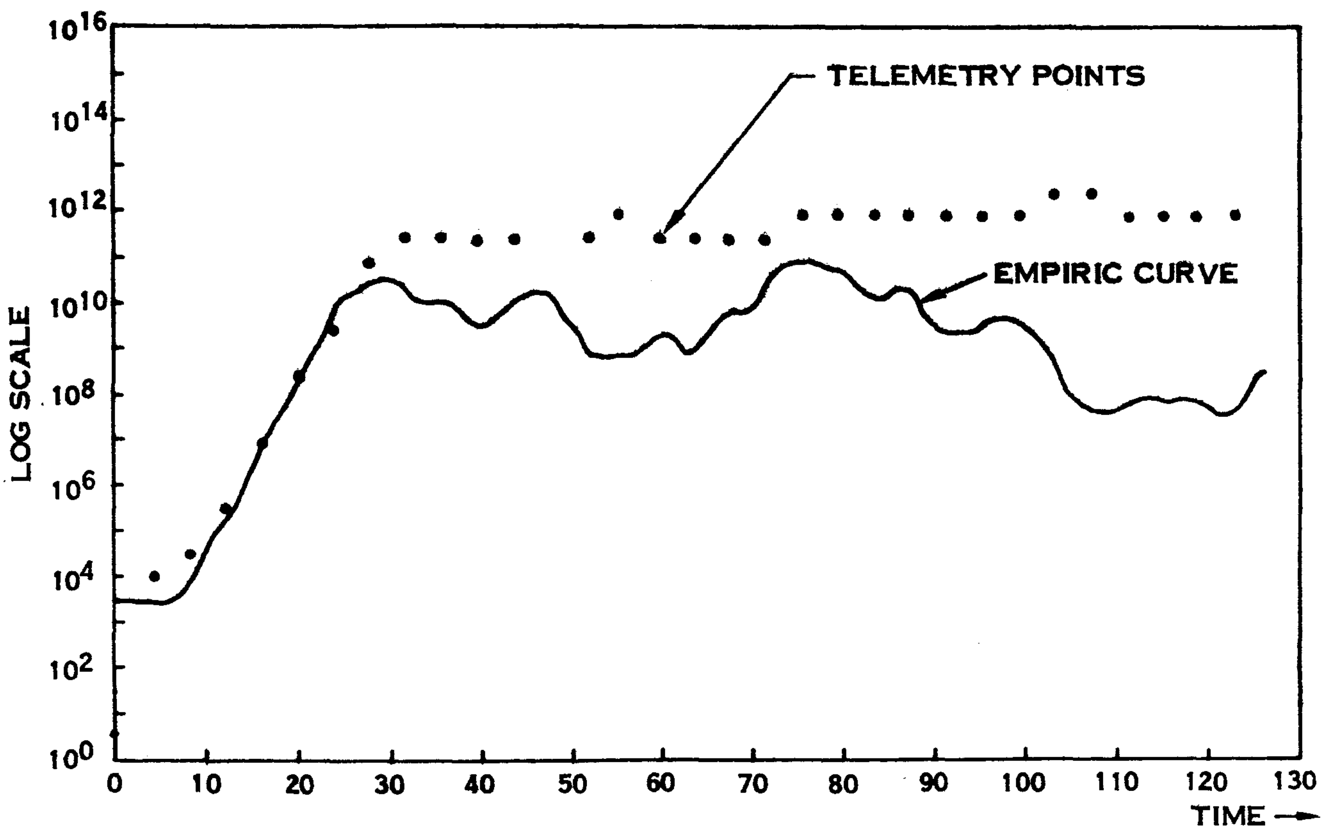
TELEMETERED OBSERVATIONS AND EMPIRIC GROWTH CURVE
FOR NOISY CULTURE

FIGURE 2

INPUTS
IDEAL CURVE 1
$r = 0$
$g = 1$
$H = .25$

DETECTOR SENSITIVITY .5 LOG UNIT
DETECTOR RANGE $10^{.5}$ TO 10^{14}

32 TELEMETERED POINTS



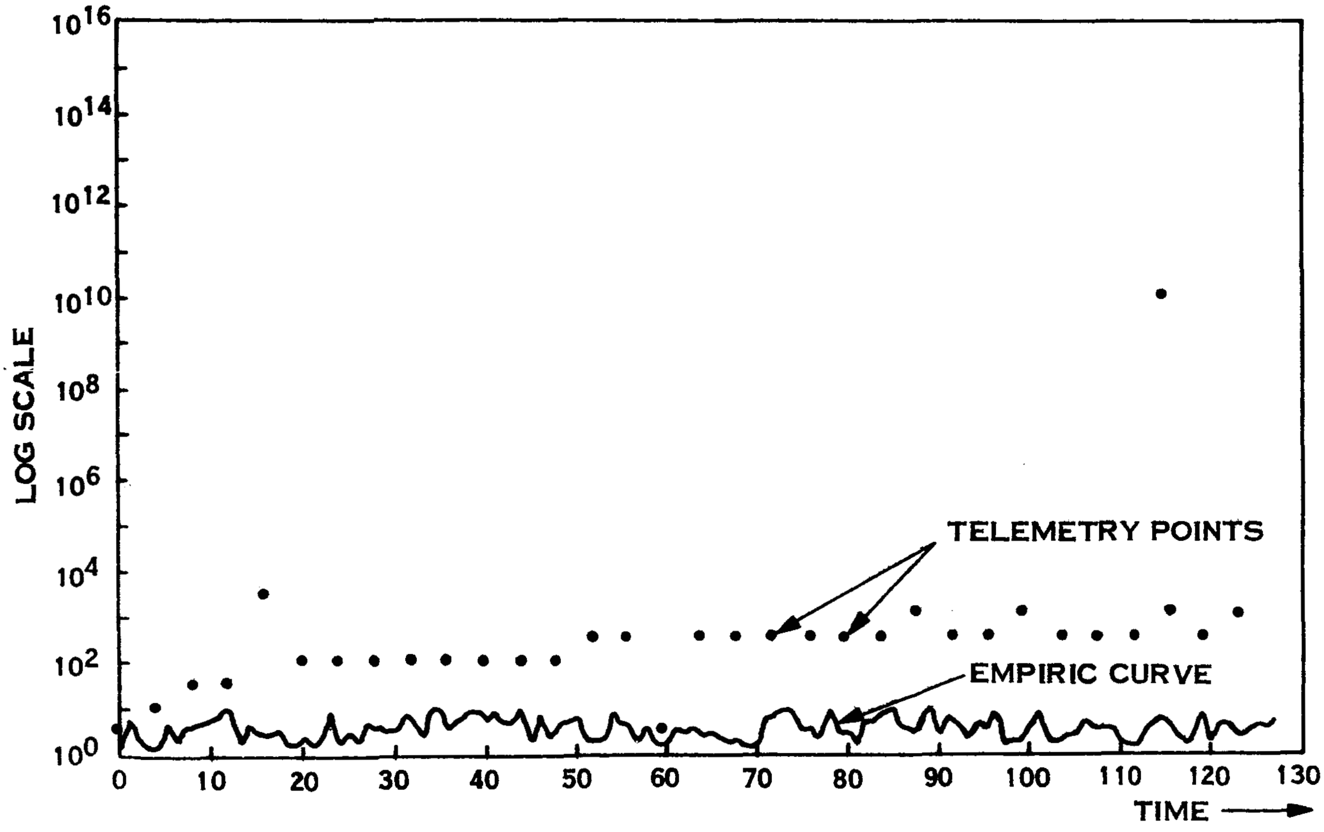
TELEMETERED OBSERVATIONS AND EMPIRIC GROWTH CURVE
FOR NOISY CULTURE

FIGURE 3

INPUT

DETECTOR SENSITIVITY .5 LOG UNIT

DETECTOR RANGE $10^{.5}$ TO 10^{14}



TELEMETRY POINTS AND EMPIRIC CURVE CHARACTERISTIC
OF DEGENERATE GROWTH OR CONFOUNDING

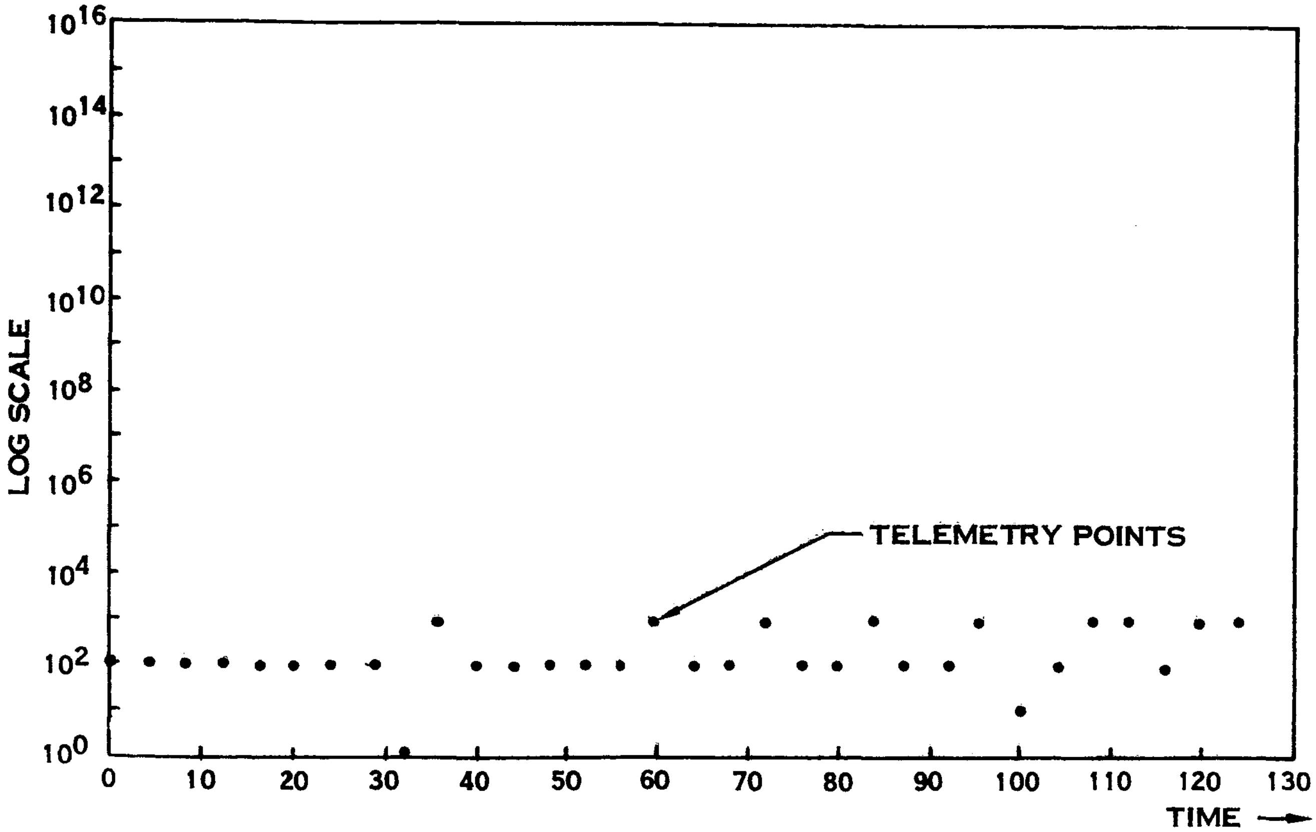
FIGURE 4

INPUT

DETECTOR SENSITIVITY 1 LOG UNIT

DETECTOR RANGE 10^2 TO 10^{11}

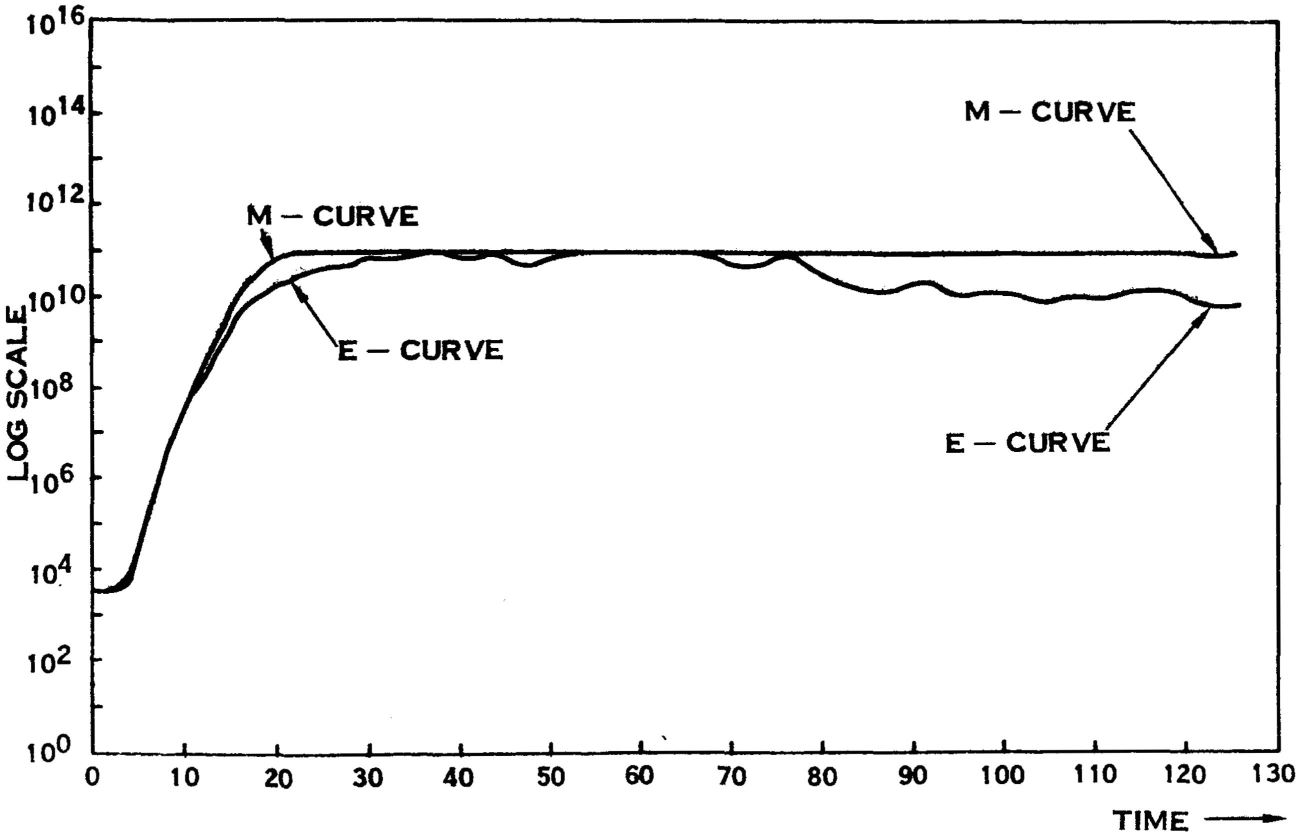
32 TELEMETERED POINTS



TELEMETERED POINTS FROM CONFOUNDING OR DEGENERATE GROWTH
USING MEDIUM SENSITIVITY AND RANGE DETECTOR

FIGURE 5

INPUTS:
IDEAL CURVE 1
r = 0
g = 0
H = .125



MODIFIED IDEAL CURVE AND EMPIRIC GROWTH CURVE FOR LARGE INNOCULUM, RAPID GROWTH CASE

FIGURE 7

INPUT

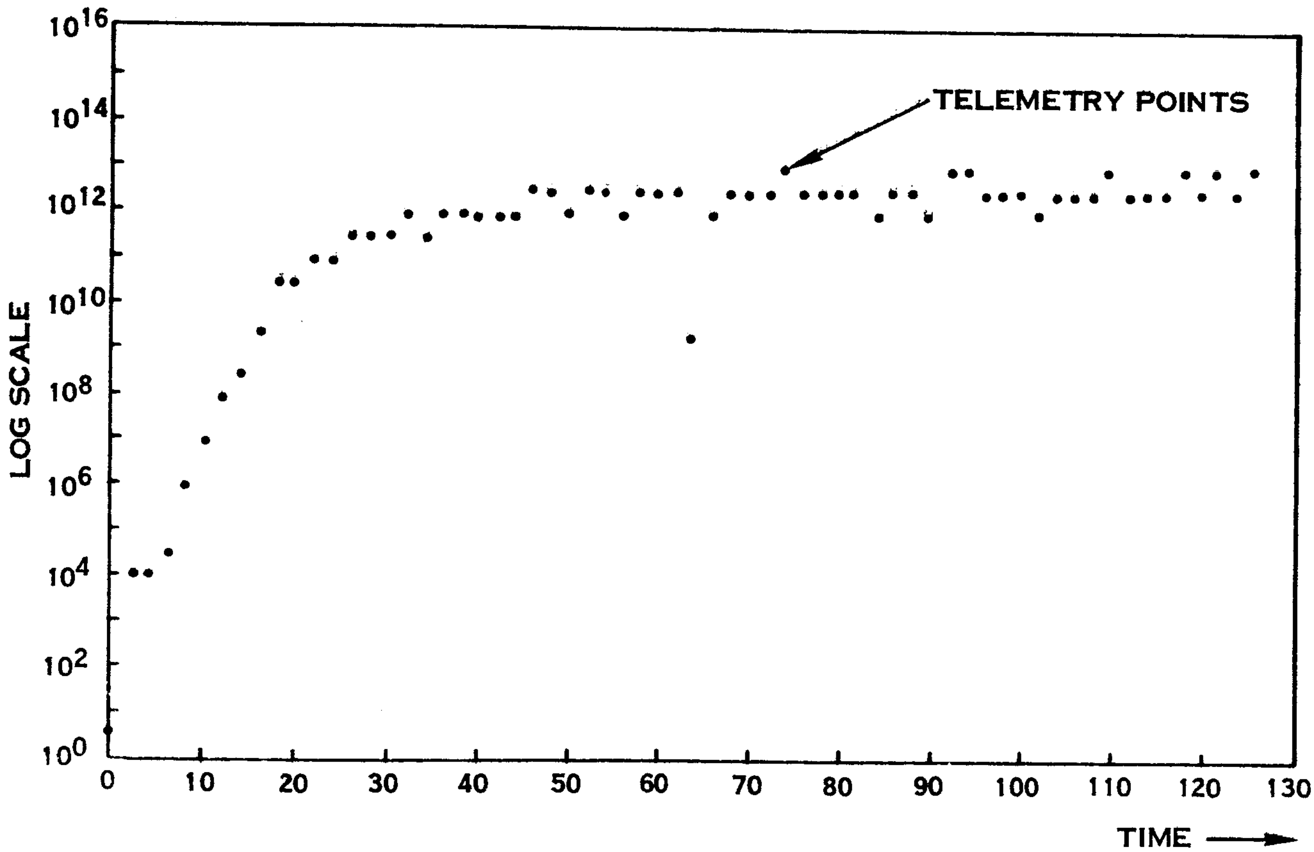
IDEAL CURVE 1

$r = 0$

$g = 0$

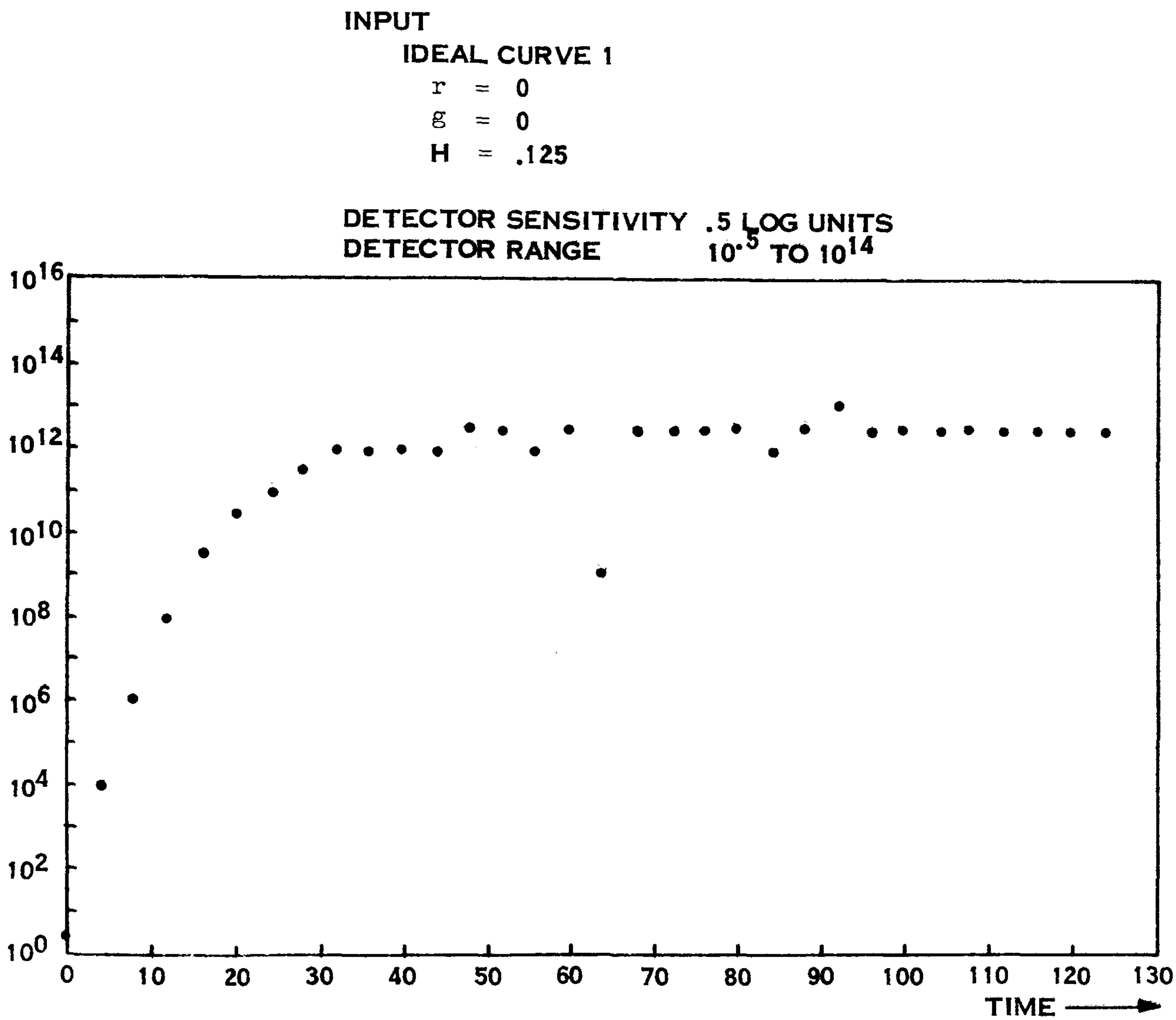
$H = .125$

DETECTOR SENSITIVITY	.5 LOG UNIT
DETECTOR RANGE	$10^{.5}$ TO 10^{14}



TELEMETERED POINTS CHARACTERISTIC OF HIGH SENSITIVITY WIDE RANGE DETECTOR WITH HIGH SAMPLING FREQUENCY

FIGURE 8



TELEMETERED POINTS CHARACTERISTIC OF HIGH SENSITIVITY
WIDE RANGE DETECTOR WITH MEDIUM SAMPLING FREQUENCY

FIGURE 9

INPUT

IDEAL CURVE 4

$r = 0$

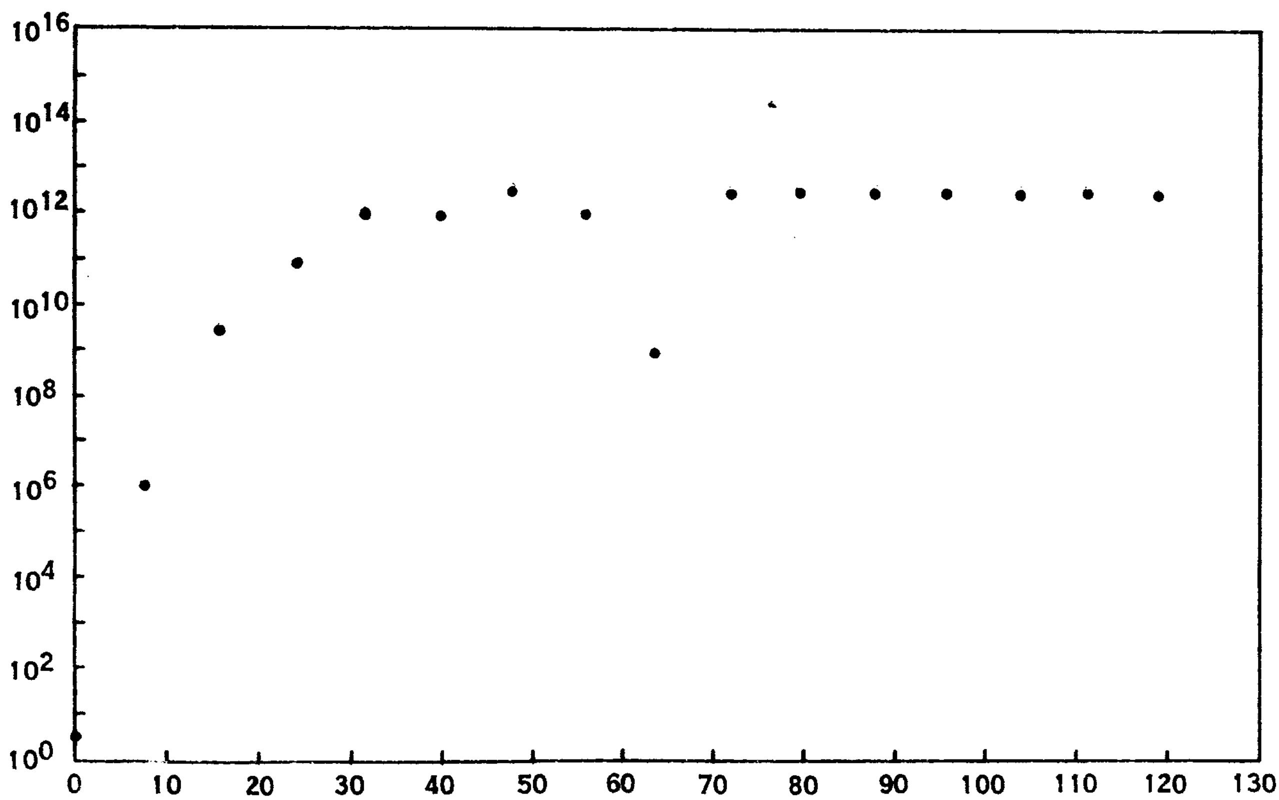
$g = 0$

$H = .125$

DETECTOR SENSITIVITY .5 LOG UNIT

DETECTOR RANGE $10^{.5}$ TO 10^{14}

16 POINTS PLOTTED



TELEMETERED POINTS CHARACTERISTIC OF HIGH SENSITIVITY WIDE RANGE DETECTOR WITH LOW SAMPLING FREQUENCY

FIGURE 10

INPUTS

IDEAL CURVE 1

$r = 0$

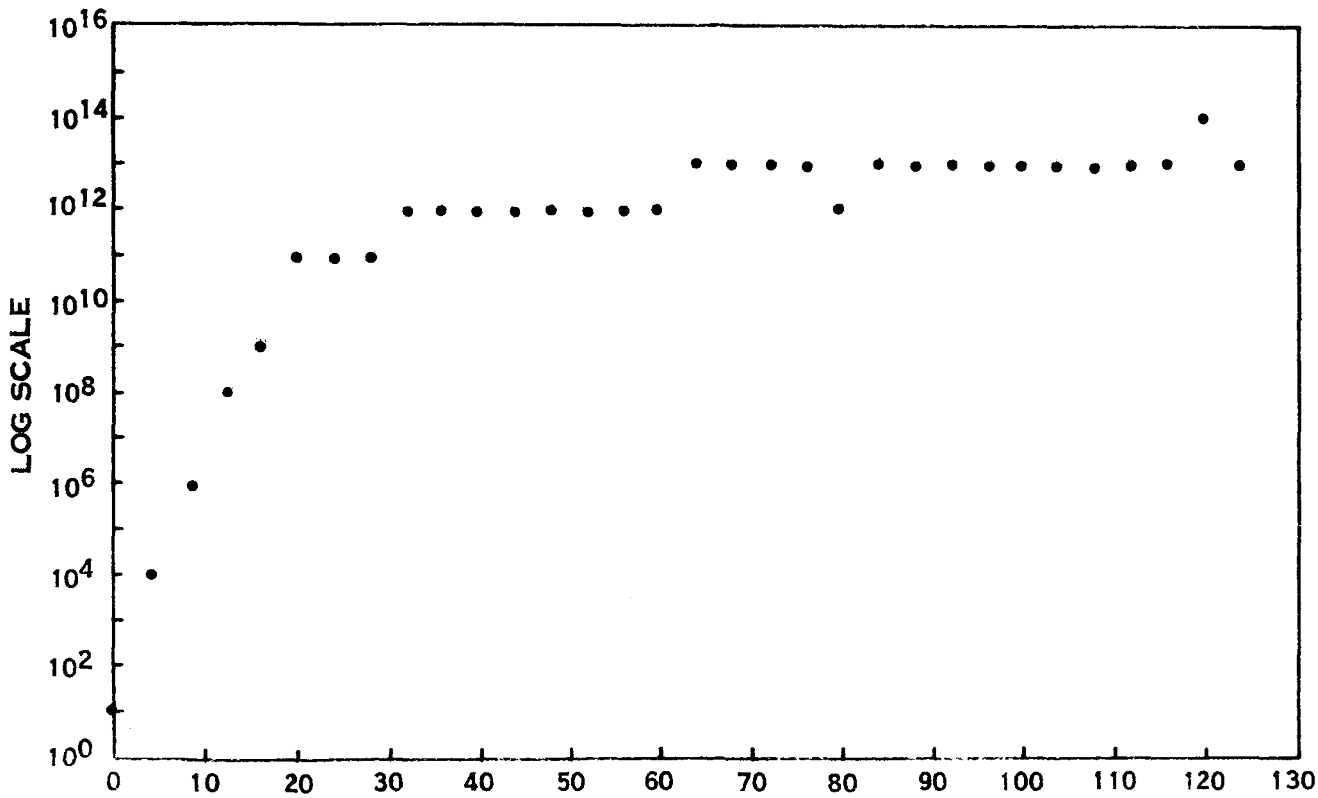
$g = 0$

$H = .125$

DETECTOR SENSITIVITY 1 LOG UNIT

DETECTOR RANGE 10^1 TO 10^{14}

32 TELEMETERED POINTS PLOTTED



TELEMETERED POINTS RESULTING FROM WIDE RANGE AND MEDIUM SENSITIVITY DETECTOR

FIGURE 11

INPUTS

IDEAL CURVE 1

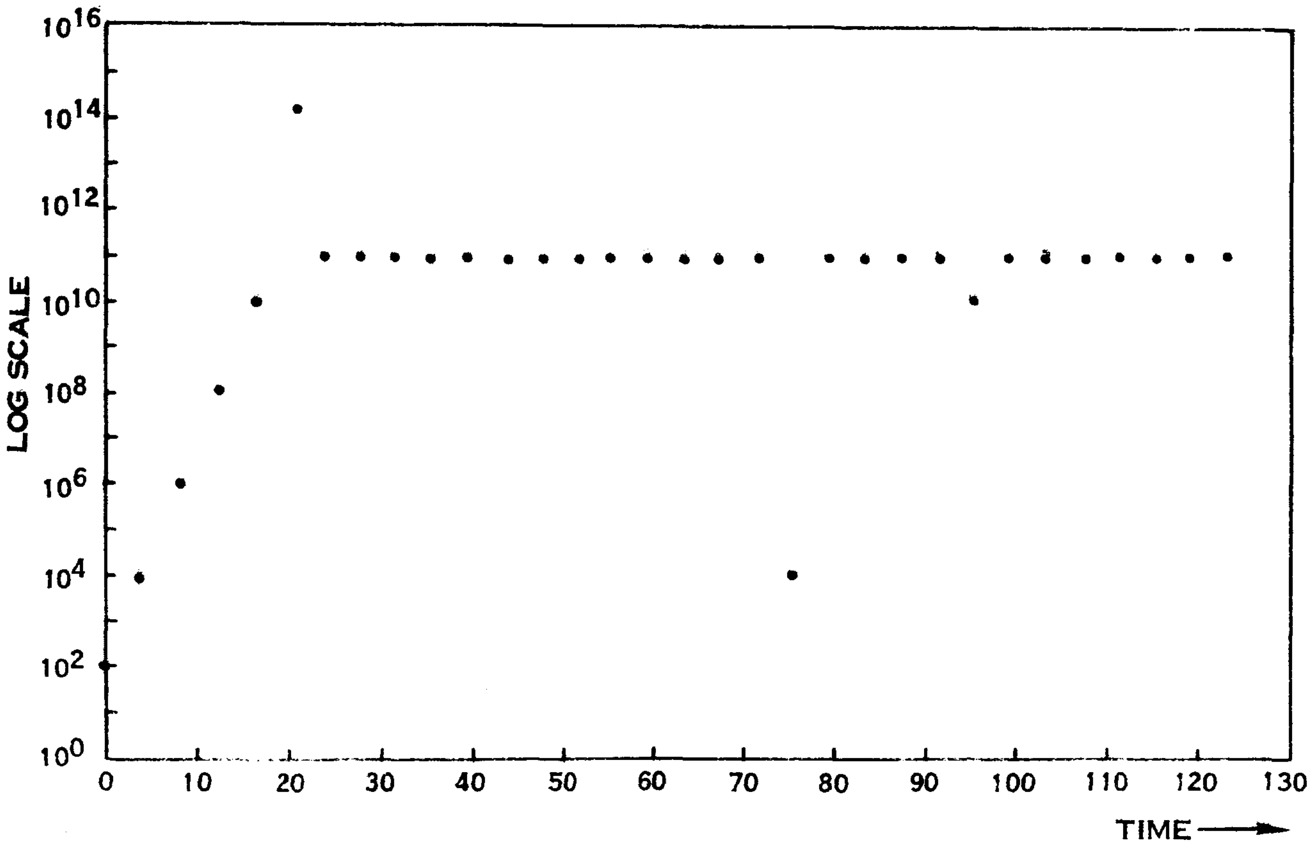
$r = 0$

$g = 0$

$H = .125$

DETECTOR SENSITIVITY 1 LOG UNIT

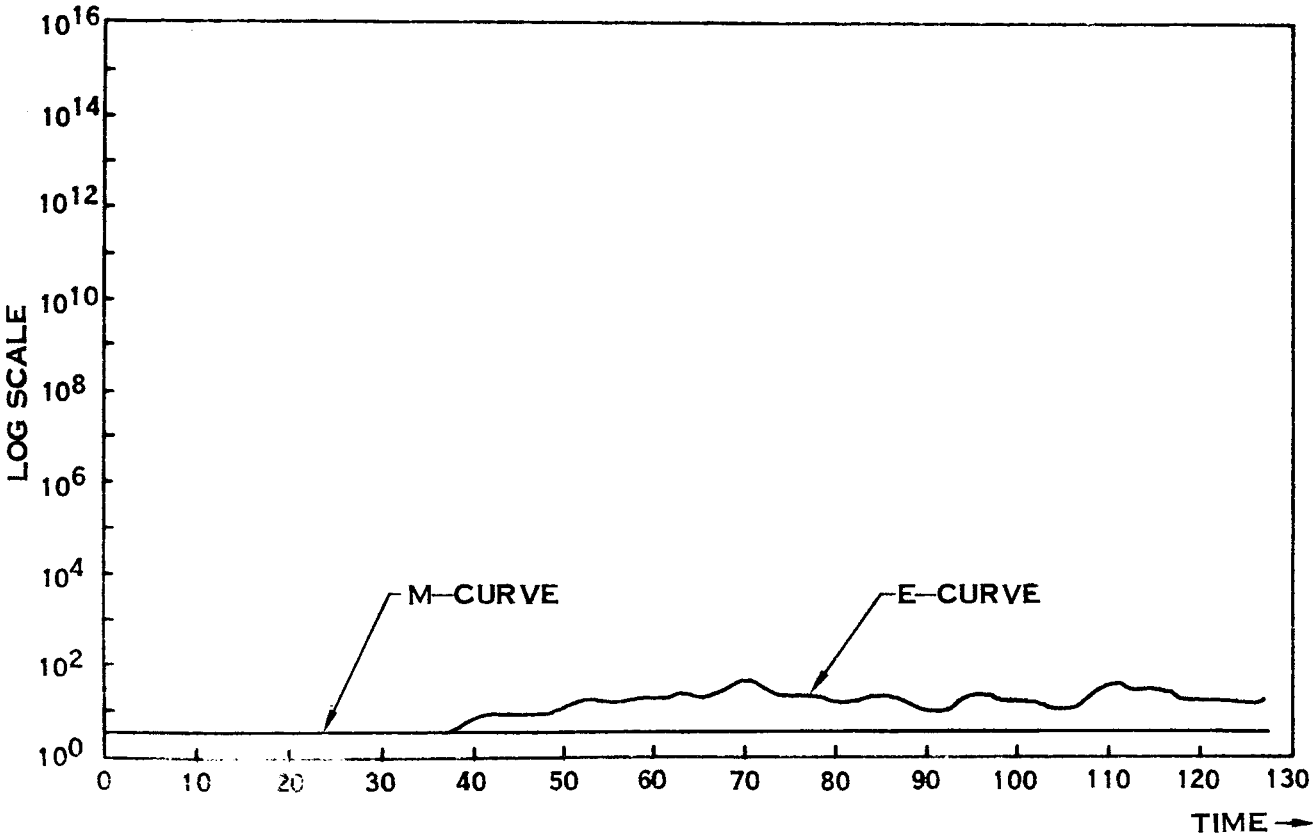
DETECTOR RANGE 10^1 TO 10^{11}



TELEMETERED POINTS RESULTING FROM MEDIUM SENSITIVITY
LOW SATURATION LEVEL DETECTOR

FIGURE 12

INPUTS
IDEAL CURVE 3
r = 9
g = 4
H = .125



MODIFIED IDEAL CURVE AND EMPIRIC GROWTH CURVE FOR SMALL INNOCULUM
SLOW GROWTH "NON-ADAPTED" CASE

FIGURE 13

INPUTS

IDEAL CURVE 3

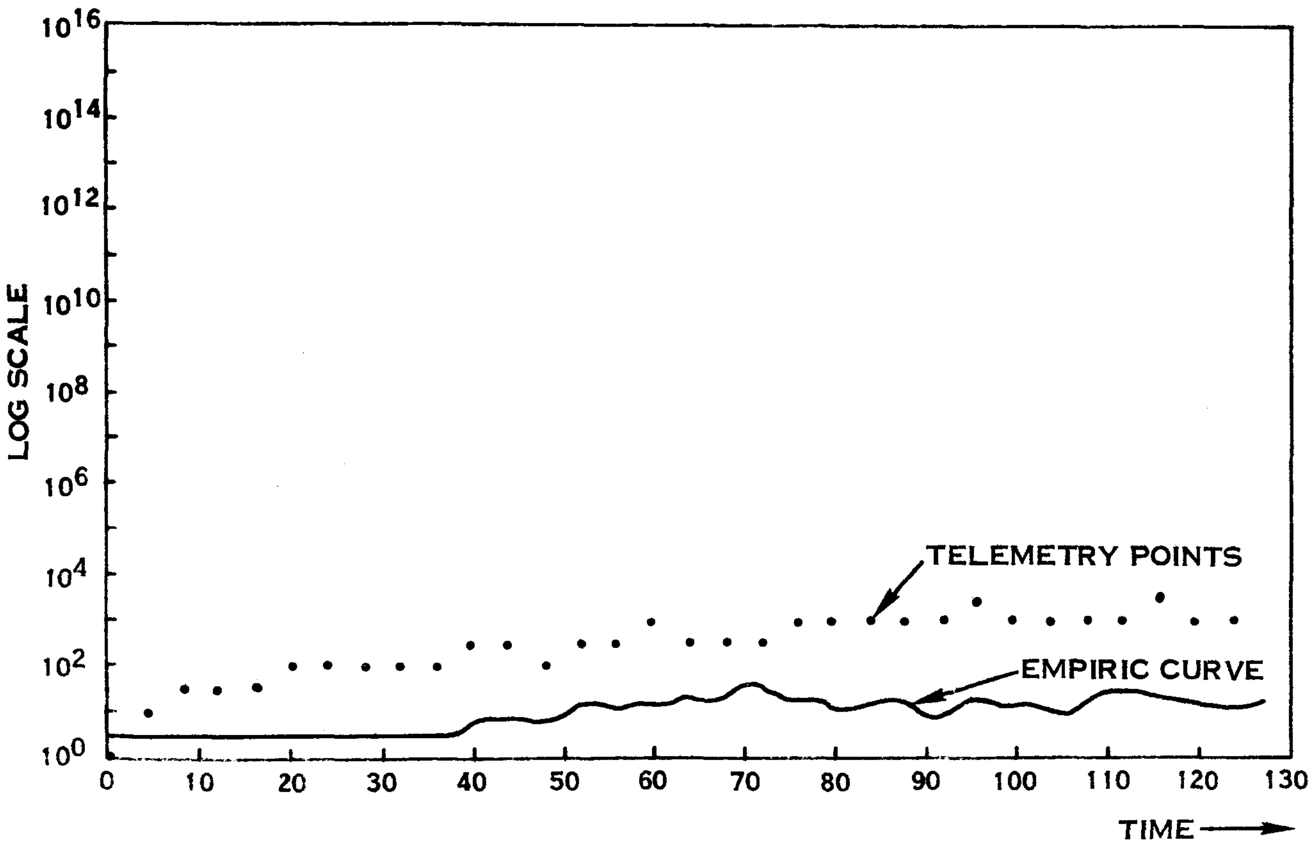
r = 9

g = 4

H = .125

DETECTOR SENSITIVITY = .5 LOG UNIT

DETECTOR RANGE $10^{.5}$ TO 10^{14}



TELEMETERED POINTS AND EMPIRIC GROWTH CURVE FOR SMALL INNOCULUM SLOW GROWTH CASE WITH HIGH SENSITIVITY DETECTOR

FIGURE 14

INPUTS

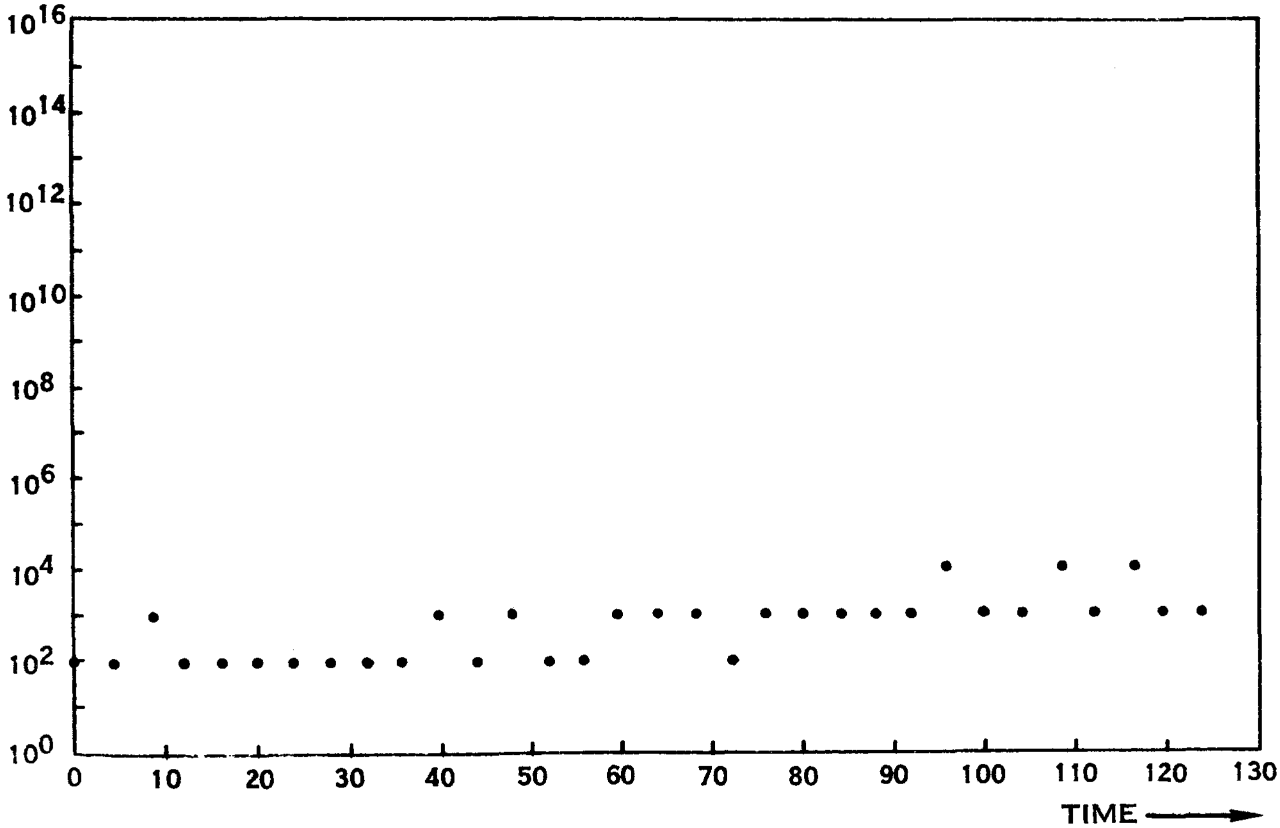
IDEAL CURVE 3

$r = 9$

$g = 4$

$H = .125$

DETECTOR SENSITIVITY	1 LOG UNIT
DETECTOR RANGE	10^2 TO 10^{14}



TELEMETERED POINTS FROM MEDIUM RANGE AND SENSITIVITY DETECTOR, SAME EMPIRIC CURVE AS FIGURES 13 AND 14

FIGURE 15

INPUTS

IDEAL CURVE 4

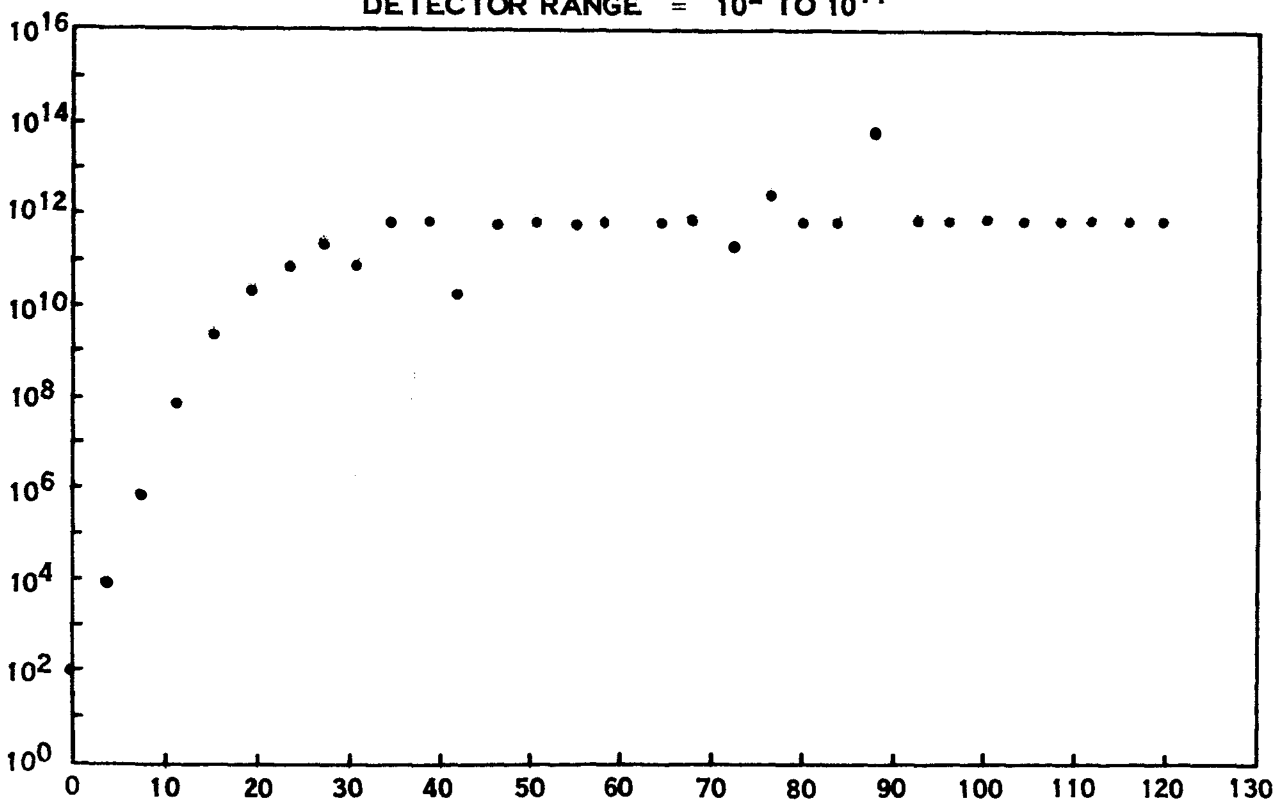
$r = 0$

$g = 0$

$H = .125$

DETECTOR SENSITIVITY = .5 LOG UNIT

DETECTOR RANGE = 10^2 TO 10^{11}



TELEMETERED POINTS RESULTING FROM NARROW RANGE
HIGH SENSITIVITY DETECTOR EMPIRIC CURVE
SHOWN IN FIGURE 7

FIGURE 16

5.0 BANDWIDTH REQUIREMENTS

Data transmission requirements constitute one of the several criteria which can be used in the selection of candidate Martian life detection experiments. The amount of intelligence which must be transmitted in order to permit adequate interpretation of the experiment varies among the several classes of experiments. Since each bit of transmitted information can be directly converted into mass of power supply and data transmission equipment, the data requirements of a given experiment is an important parameter of total lifted payload.

Communication theory sets bounds on the rate at which information can be transmitted over a channel of arbitrary noise-content subject to a given mean square (or other) criterion of fidelity. In addition, communication theory establishes the existence of an asymptoticaly most-efficient-encoding for any message source of arbitrary structure but, in general, does not present constructive formulae for constructing such encoding for all cases.

In order to construct an optimum, or near optimum, encoding of a message source it is necessary to first specify what features or parameters of the original signal must be transmitted in order to convey the desired intelligence to the receiver. Following this step and using the presumed noise content of the channel, the required average rate of transmission, and information regarding source of noise or bias that may be contributed by the instruments employed in conducting the experiment, a suitable communication system can be designed and which will incorporate the desired level of redundancy.

A preliminary study was made on the several classes of candidate Martian life detection experiments during this reporting period in order to obtain insight into the types of signal parameterization required for adequate interpretation of the experiment. Knowledge of such parameterization would serve as a necessary first step in devising efficient codes. This study, however, did not extend to detailed specification of such codes, since the parameters of the physical communication equipments of the mission are as yet unspecified.

5.1 Type A Experiments

Increase in cellular population over time. These are microbial growth studies in which, ideally, we test the hypothesis that:

$$\ln N = \ln N_0 + \beta t \qquad (1)$$

where $\ln N$ is the natural logarithm of N, the number of cells per unit volume at time t. The objective of interest is value of β, the growth rate of the organisms. The figure of merit is $\hat{V}(b)$ the sample variance of b which is an estimate of β computed from the data. $\hat{V}(b)$ is of course also computed from the data so that it, too, is an estimate of $V(b)$ the "true" variance.

Since (it is hoped) that the investigator can specify in advance the magnitude of $V(b)$ he is prepared to tolerate we take the latter as a design parameter.

Let time t from the start of the experiment be divided into intervals Δt. Then owing to the lack of perfect control, etc., in the microbial propagator, the number of organisms at the end of the i th unit of Δt

will be:

$$Z_i = \ln N_i = \ln N_o + t_i\beta + e_{3i} \qquad (2)$$

where e_{i3} is a random biological variate with zero expectation and variance $E(e^2_{i3}) = \sigma_3^2$. Z_i is now subjected to electronic noise e_{i1} and digitization noise e_{i2} with:

$$E(e_{i1}) = 0 \quad E(e_i^2 1) = \sigma^2$$

$$E(e_{i2}) = E(e_{i\,2}^2) - E(e_{i2})^2 = \Delta^2/12$$

where Δ is the interval between digitization levels.

These two noise sources are added at each transmission of Z_i to yield:

$$X_i = Z_i + e_{i1} + e_{i2} \qquad (3)$$

which is the quantity actually transmitted to earth Z_i.

It can be shown that:

$$b = \frac{Xn - Xo}{N\ t} \qquad (4)$$

i.e., the last less the initial received datum divided by the total time, is the best estimate b of β.

The expected variance of b is:

$$E[V(b)] = \frac{1}{N}\sigma_3^2 + 2\sigma_2^2 + \Delta^2/6 \qquad (5)$$

The quantity Δ can be equated to C_2^{-T} where C is the full scale output of the detection device and r is the number of bits per datum. If T is the total number of data bits available during the lifetime of an experiment, then expression (5) is minimized by:

$$N = \frac{0.6T}{\log_{10}(0.231C^2\ T/\sigma_3^2)} \qquad (6)$$

If the biological variance, σ_3^2 is scaled in terms of the parameter V where:

$$\sigma_3^2 = 2^{-V}C^2T(0.231)$$

then $r = V/2$ and $N = 2T/V$.

In general, this class of experiment (as suggested by the SOUP GEDANKEN simulation) requires moderately large numbers of equispaced observations with relatively small numbers of bits per observation.

In several variations of this class of experiment, it is to be anticipated that particles of Martian material will be introduced into the culture chamber and will thus interfer or be aliassed with the production of cellular material. This condition will require that the local minimum be detected of a time function of the type:

$$X = A\,(e^{\beta t} + e^{-kt})$$

where β is the growth rate of the organisms and K is the idealized settling rate of the particles.

The precision with which the minimum can be detected depends on the relative magnitude of β and K and on Δt the sampling interval.

Reasonable estimates of β and κ can be used to estimate reasonable values of Δt which in turn will provide bounds on N, the number of sample points required and thus on T the total number of bits which must be provided. Experiments of this class are not demanding in total number of bits required.

5.2 <u>Type B Experiments</u>

In this class of experiment the presence of viable organisms is detected by their ability to convert a labeled substrate to a metabolic product. Generally the total amount of metabole liberated up to time t is measured.

Since the total number of organisms is changing during the experiment due to a birth and death process, the rates at which product is liberated change as functions of the changes in size of the viable population.

The general form of the response of the system can be represented by an integral equation of the form:

$$y(t) = \int_0^t F(t-\lambda)\, R(\lambda)\, d\lambda + R(0)$$

where F (θ) is a "metabolizing" function and R (θ) is a substrate "supply" function. Although a good deal of theoretical work on the form of these functions has been carried out both here and elsewhere, the unavailability of precise information on the form of these functions makes detailed analysis impossible.

Empirical observations on systems of this class suggest that functions of the type where A is a function of the initial biologic inoculum are moderately good approximations of the observed response of the system.

In general, it can be shown that the same considerations applying to the Type A experiments, namely, frequent sampling with a relatively small number of intelligent bits per sample apply equally well to the present class of experiment.

5.3 Type C Experiment

In this class of experiment intelligence is contained in a signal which varies in time in such a fashion as to describe a series of peaks of different widths, amplitudes and shapes. The example which has been studied is the gas chromatograph, the detector output of which consists of a series of Gaussian shaped peaks if the column material is appropriately designed so that it effectively separates all components of the sampled mixture. If effective separation of all components is not achieved, then peaks will not be well separated; some may be skewed, and some will be superimposed on one another so as to appear assymetrical, and the baseline from which the well separated peaks emerge will be seen to shift in a complicated fashion and may appear to have a number of small jagged variations imposed on it.

We may usefully distinguish between those components which the column, the detector, and the related sample injection and temperature control regimen are specifically designed to detect, and the residue of materials in the sample which are not well separated. The peaks corresponding to the former are Gaussian in shape and the width of each peak is very small relative to its amplitude. The presence of the components which are not well separated is revealed either as skewed or lumpy "peaks" with wide base and ill-defined maxima, or as smooth or jagged variations in the base line which may be thought of as confounding effects of components in which there is little interest.

If the experimenter wishes only to recover information regarding the presence of the components for whose detection the instrument is well adapted, then a number of alternative coding procedures with differing bandwidth requirements are available. If these peaks are Gaussian in shape, and if $4\alpha T$, $\alpha < 1$ is the base width of the narrowest peak of interest, then the Fourier spectrum of the signal

$$F(w) = A \exp(-1/2 \; {}^{2} T^{2} w^{2}),$$

w in radians/second, T = total length of the time axis and A is an arbitrary constant.

A "naive" coding procedure which samples the amplitude of the detector output signal at equal intervals of time requires $31.4\alpha T$ samples/second to recover 95% of this spectrum. This naive coding leads to inordinately high bit requirements and results in what might be considered excessively good resolution. For example, a total record of 30 minutes with $4\alpha T = 1$ second requires 2.26×10^{5} samples or 125.5 per second. If it is desired to measure the maximum amplitude with an accuracy of approximately ± 1.5% then 6 bits will be required to transmit each sample, and a total of approximately 1.25×10^{6} bits will be required. Such a coding scheme will resolve almost perfectly all peaks of base width 1 second or more, and will also describe the behavior of the base line; it will do so no matter how many peaks occur during the 30 minute period, provided they are separated by a distance of at least one second.

The level of resolution obtained is unnecessarily high, and experience suggests that an equal interval sampling scheme involving no more than 5 samples per second is likely to be adequate. Such a naive coding procedure requires 54 x 10^3 bits to transmit a 30 minute chromatogram no matter how many peaks it contains.

This naive coding scheme is highly inefficient if the number of peaks about which data is desired is small. In such cases it is much more efficient to devise a code which uses the available bits to transmit information only about the behavior of the waveform in those regions where interesting peaks occur, and to transmit no information about the behavior of the detector output at other times. Several peak detection and encoding schemes have been investigated. All are similiar in that the number of bits required to transmit data regarding detected peaks varies with the number of peaks occurring and is very low for the expected number of peaks. Thus all require far fewer bits than the naive scheme, and provide equal or better resolution where resolution is expressed as function of the width of the narrowest peak detected. They differ in respect to the fidelity with which the original detector output can be reconstructed from transmitted messages, the kind and number of peaks about which information is transmitted, and the penalties resulting from corruption of those messages due to communication noise.

These schemes are briefly described below:

1. Peak amplitude and time. This system requires a differentiator which calculates the slope of the detector output signal, and a clock which measures elapsed time in suitably small increments. The output of the detector is positive and the output of the differentiator is zero whenever the detector output reaches a local maximum. At these times a message corresponding to the amplitude of the detector output (peak height) and the elapsed clock time (retention time) is transmitted. The number of bits required to describe n peaks is $n(j + k)$ where there are 2^j amplitude intervals measured and the total duration of the experiment is divided into 2^k time increments. Thus, if $j=6$ and $k=11$ (which will be the case if the experiment lasts 30 minutes and time resolution to the nearest second is desired) then only 4250 bits will suffice to describe a chromatogram of 250 peaks.

The disadvantage of this system is that all local maxima will result in the transmission of a message indicating that a peak has occurred, so that slow shifts in the baseline corresponding to confounding due to components which are not well separated, and wide, ill-defined peaks cannot be distinguished from the narrow well-separated peaks of interest. These, however, can be readily filtered out by several techniques such as discriminating among maxima as a function of one or more of the following:

(a) The slope of the curve immediately preceding and following the maximum, so as to eliminate those with gradual slopes.

(b) The value of the peak amplitude, so as to eliminate peaks close to the true baseline.

(c) The amount by which the detected local maximum differs from the value of the preceding minimum (where the minima are detected as alternate zero outputs from the differentiator) so as to eliminate the small jagged wiggles and peaks frequently superimposed on other peaks or on the base-line.

(d) Peak width as determined by the time elapsing between the two minima adjacent to the peak, so as to eliminate those which are too wide or too narrow to be of interest.

If discrimination procedures such as these are employed, then peaks not filtered out can be easily reconstructed from the received data, since every message describing a peak can safely be interpreted as corresponding to one of a small class of "well separated" peaks of known shape. This will also result in further reductions in the number of bits which must be employed, since if only 50 of the 250 peaks have the desired shape, only 850 bits will be required to describe these. Coding schemes very similar to that described here are favored by many experimenters.

Although this scheme requires very few bits and greatly simplifies the interpretation of the experimental results, it restricts the class of compounds about which information is received to those which the GC is specifically designed to detect. If compounds of this nature are not present in the sample examined, then no information describing the behavior of the detector output signal will be received. In those cases in which the GC is to be used solely to detect identifiable components with known physical properties (and in which the experimenter has considerable advance information regarding the identity of the compounds which are likely to be present in the sample) the restrictions resulting from such a coding procedure are not undesirable. A good example of such an experiment is that of the analysis of the atmosphere of Mars; specific GC's with restrictive data encoding schemes are acceptable because there is considerable advance information about the identity of the components which are likely to be present, and the GC can be designed to respond specifically to these.

However, if we consider possible applications of GC techniques to situations in which there is much less advance information about the possible constituents of the sample, it becomes obvious that a restrictive coding scheme may be less desirable in that it limits the number of compounds about which information is received to those which the GC is specifically designed to separate and detect. In such situations a GC designed to look only for a limited class of components cannot be so easily justified, and there may be a need for data encoding schemes which are capable of providing information about the occurrence of components which are not well separated, for use with "general purpose" GC's, or to decrease somewhat the selectivity of special purpose GC's. Several such schemes were postulated and compared with the restrictive coding scheme described above. Only one will be described here since it has not been possible to evaluate the differential advantages of the several schemes, and all of the others are variations of this procedure.

2. Local minima, maxima and zero level detector. This scheme employs the same differentiator and clock as that described above, but employs simple logic to examine the signal amplitude at every occurrence of a local minimum and maximum, and to distinguish the times at which the signal arrives at or departs from the true baseline from maxima and minima occurring above

the baseline. The messages transmitted are:

(a) Signal amplitude and clock time at every time the value of the differentiator output changes sign and the signal amplitude is not zero. (These are local minima and maxima).

(b) Clock time alone at every time the differentiator output departs from zero or becomes zero, if at that time the value of the detector signal is zero. (This corresponds to the times at which the signal departs from and declines to the true baseline).

This procedure will permit the experimenter to distinguish several peaks occurring close together in time even if the signal does not decline to the baseline between them. Thus, tall narrow peaks corresponding to "well separated" components can be readily distinguished from ones which are wide or skewed, or which correspond to large numbers of components which are not well separated. The number of bits required to transmit each peak depends upon whether they are superimposed on the true baseline. If so, $3k + j$ bits will suffice. If the peaks are separated by non-zero minima, then $2(j+k)$ bits will be required for each. Since this system will provide much more information about the behavior of the output signal than will the preceding one, it will in general require far more bits. For example, if half the peaks require $2(j+k)$ bits and half require $3k + j$ bits, then a 250 peak chromatogram will require 9250 bits for its transmission, as opposed to the 850 required for the more restrictive system which filters out all but 50 of the peaks. However, the resulting representations of the signal contain considerable redundancy due to the fact that minima and maxima must alternate, and departures from the zero baseline must be preceeded by arrivals at it. As a result, it is less readily corrupted by communication noise.

It will be noted that this procedure may have some use in identifying bias or machine error of other kinds. Its primary utility, however, lies in the fact that it reduces the specificity of the chromatograph experiment at a relatively slight cost in bandwidth, while retaining the advantage inherent in the fact that the shapes of all "well separated" peaks of interest can be easily reconstructed from the received data. It should be pointed out that the restrictive coding scheme requires equally complex logical circuitry if good peaks are to be filtered from bad ones, and if the height of these peaks relative to the shifting "baseline" is to be distinguished. The advantages accruing from reduced specificity of the overall experiment have not been evaluated.

6.0 PROBLEMS OF THE DESIGN AND MANAGEMENT OF SPACECRAFT STERILIZATION PROCEDURES.

During the past year preliminary studies were made on the general problem of spacecraft sterilization. These studies were designed to elucidate (1) the nature of the nominal level of spacecraft and mission sterility specificed by national policy, (2) the cost of attaining a specified level of component sterility in terms of the reliability of the component, (3) the assay procedures and quality control techniques to be employed to insure that a given class of component actually complies with the nominal or target level of sterility specified for it.

For a large class of components, attainment of a given level of sterility will be achieved by imposition of process control accompanied by sampling followed by (possibly) destructive microbiological assay. Quality control tests based on microbiological assay are subject to error variances considerably greater than those associated with conventional quality control tests by variates. In situations of this class, there exists a question of how to determine near optimal assignments of the total quality control budget to a fraction, f, assigned to monitoring the process and a fraction (1-f), to sampling and testing the product itself. This situation is the classic decision problem as applied to quality control but with the added dimension of diminution in component reliability as a consequence of sterilization.

The relationship between level of sterilization and the reliability of a component may be symbolized as:

$$R_f = T_1(R_i)$$
$$C_f = T_2(C_i)$$

where:

R_f = reliability after sterilization
R_i = reliability before sterilization
C_f = contamination after sterilization
C_i = contamination before sterilization

The symbols T_1 and T_2 denote particular transformations of the initial states of the components. The structure of T_1, T_2 reflects the particular sterilization procedure used. Clearly, the ideal procedure is one in which:

$$R_f = R_i$$
$$C_f = T_2(C_i)$$ i.e., it is one which does not affect reliability.

The structure of T_1, T_2 is obtained empirically from experimental data. These experiments entail errors, however, of the form $e_1/\sqrt{c_1}$, $e_2/\sqrt{c_2}$ where c_1, c_2 are the costs of the experiments. Since e_2 is a biological variable, in general we may assume $e_1 \ll e_2$ and $c_2 \sim c_1$. C_i is the initial microbiological contamination of the component and its level determines in part the nature of the T_2 which must be applied to reduce C_i to C_f. The quantities C_i and C_f, however, can only be estimated by microbiological assay which entail error of $e_2/\sqrt{c_3}$, $e_2/\sqrt{c_4}$ where e_2 is, again the intrinsic error of the assay and c_3, c_4 are the assay costs.

If C_{f0} is the quality level assigned to the component we are required to select a T_2 such that the probability that $C_f < C_{f0}$ is not less than some quantity $1-\alpha$.

To meet this constraint we must utilize assay data on C_i and select a T_2 such that due allowance is made for the errors in constructing T_2 <u>and</u> the errors in assessing C_i. In general, large errors in T_2 and in the assay lead to the selection of T_2's which must include large safety factors.

"Rigorous" T_2's however will, in general, tend to reduce R_f or increase the cost of components with average $R_f = R_{fo}$.

During the course of any future efforts we plan to further explore the usefulness of general models of sterile spacecraft fabrication in order to determine the areas in which detailed analysis offer good prospects in improving operational efficiency.

7.0 ANALYSIS OF EVALUATION PROBLEMS ASSOCIATED WITH VISUAL RECONNAISSANCE EXPERIMENTS

7.1 Introduction

Life detection by means of photographic surveys on a macroscopic or microscopic scale has frequently been proposed, and these proposals have great appeal even though it is recognized their cost is enormous when considered from the point of view of the bandwidth required to transmit photographic data. The chief appeal of photographic experiments lies in the fact that visual evidence is the most compelling evidence normally utilized in confirming the presence of organisms, and it is possible to conceive of many kinds of things that might be portrayed in a photograph and which would constitute entirely convincing evidence that life exists on Mars. Considered on the basis of the bandwidth required to transmit it, one such picture might be well worth the cost. The liklihood that such a photo will be obtained is much more difficult to estimate, and consequently visual reconnaissance experiments are very difficult to evaluate.

The need to attempt some such evaluation derives from the fact that transmitting visual data requires exceedingly large communication capability. For example, the power used to transmit a medium quality photograph by means of the sterilized battery-powered direct communication systems considered for a Mariner-type lander corresponds roughly to a minimum of 100 to 150 lbs of battery. Estimates are not firm. A medium quality photo consisting of a 500 by 500 array of points each of which may take on one of 8 brightness levels requires $.75 \times 10^6$ bits to transmit the brightness information alone and an additional .25 to .75 x 10^6 bits will be required to provide limited registration data. for a total of 1 to 1.5 x 10^6 bits.* Each pound of battery may be somewhat optimistically supposed to permit the transmission of 10,000 bits of this data. It is obvious that even if power is not supplied by batteries and the weight penalty per bit is much lower, photographic data will still be relatively costly. Consequently a brief study was undertaken to identify some of the factors which should be taken into consideration in evaluating the worth of visual reconnaissance experiments and the terms in which such an evaluation might best be expressed.

The problem of evaluating visual experiments is one of estimating the liklihood that the transmitted photos will be recognizable photos of recognizable objects of biological significance as a function of (a) the location and manner in which the photographs are obtained and (b) the bits allocated to their transmission (number, size, contrast levels and grain of the photos). Three questions are relevent here: How likely is it that a biologically significant object will be encountered in the field of the camera and how does this vary with the scale and the manner of making the photo? How likely is it that an object of Martian biological significance will be recognized as such, and how does the recognizability

---

* If the transmission system were purely digital and the transmission of each picture point did not occur in some predetermined sequence then coding such a picture would require complete registration information (e.g. a message describing accurately the location of each of the 250 thousand points) or 15 x 10^6 bits. The sequential nature of the transmission system reduces the amount of registration data which must be sent, but it can be estimated that a 1.5 x 10^6 bit picture would still be slightly fuzzy.

or the reliability of a given interpretation vary with the number and detail of the photos and the manner in which they are made? Are there practical ways in which the number of bits required to transmit significant visual information can be reduced, by suitably coding each of the photos to eliminate much of the redundancy. by automatic selection techniques which select only the "best" photographs for transmission, or even by substituting a suitably programmed computer on Mars for the observing human eye? Because this last question requires a survey and evaluation of current work in visual data processing, it was not thoroughly explored during the reporting period. Consequently the results of this aspect of the study will not be reported, except to say that no proved and reliable picture coding techniques seem to be available (although some very promising approaches are under development.)

The following is a summary of the conclusions developed during the reporting period. Because microscopic and macroscopic visual reconnaissance have been proposed for somewhat different reasons, and represent somewhat different evaluation problems, these will be discussed separately.

7.2 Microscopic Visual Surveys

The objective of a microscopic visual survey is to search for microorganisms on Mars. The advantages of microscopic as opposed to macroscopic visual surveys are primarily derived from assumptions that the number of organisms per unit area or unit volume of the surface of Mars is a decreasing function of size of the organism. This seems to be a reasonable assumption in view of the fact that this is the case on earth. Another plausible argument put forward in defense of such experiments is that single celled or small organisms indigenous to Mars are more likely to be identifiable from pictures of them than are larger organisms. If these assumptions can be accepted as valid, then evaluating microscopic visual experiments is a matter of determining the probability that a good "recognizable" microphoto of a Martian organism will be obtained as a function of the following:

The number of photos transmitted

The resolution of each photo

The manner in which the photos were obtained, including resolving power of microscope, number and frequency of photos per field of view and changes in such variables as focus, phase contrast, etc.

Treatment of the sample, including use of stains, elimination of particulate matter, and other separation techniques.

A brief consideration of these factors lead to the following interpretations of their possible significance:

7.2.1 The greater the resolving power of the visual system, the more photos are necessary to inspect a given quantity of sample. If a phase contrast light microscope is used and if light intensity, phase, and focus are fixed, then many of the resulting microphots might be unrecognizable due to the fact that these settings are inappropriate to the object (whether biological or not) observed. Allowing these variables to vary in some fixed way for each field inspected, and transmitting all the resulting photos would increase the probability that at least one of these photos was appropriate to the object encountered, but at the high cost associated with the extra transmission.

7.2.2 An ordinary light microscope with less resolving power would not require varying the phase contrast or, possibly, focus, but it could not resolve organisms less than some fixed size. If the size-frequency assumption that larger organisms are less frequent is valid, then the larger organisms detectable by means of a light microscope would be less numerous and consequently proportionately more quantity of sample would have to be visually surveyed in order to insure a given level of probability of encounter with at least one organism.

7.2.3 The average number of microorganisms per unit of material examined might be increased by techniques which separate them from inorganic particulates, but evaluation of alternative separation techniques and selection of the best ones assumes more knowledge of these organisms and of the properties of the surface matter than is now available.

7.2.4 Martian microorganisms might not be clearly recognizable as such. It is not easy to distinguish between living and non-living matter even when we have a great deal of information about the sample and can control many aspects of the visual situation; we recognize things best when we know what we are looking for. Living organisms are most readily recognized because of the way that they move or because they possess clearly recognizable organelles. If moving pictures are not transmitted, then it is unlikely that movement can be employed as a recognition property. (No automatic scan techniques capable of distinguishing the distinctive movement characteristic of living organisms from other kinds of movement are now available). Organisms not possessing clearly recognizable and distinctive organelles or other component features in common with their terrestrial counterparts might be unrecognizable as such, and at present we have no grounds for assuming such similiarities. It has been claimed that trained biologists can perceive and distinguish general aspects of "regularity" which characterize all living organisms, but this claim has not been tested and the difficulties associated with interpreting the photos of "organized elements" in carbonaceous chondrites would tend to suggest that the presence of similiar "regularities" in Martian microphotos would not constitute conclusive evidence even though it might be very interesting evidence. Color contrasts obtained through the use of selective stains would aid identification, but might require addition bits in order to take full advantage of the information present in the differential response to the stains.

7.3 Macroscopic Visual Surveys

Macroscopic surveys have been recommended on the gounds that the number of photos required to scan the immediate vicinity of the lander is not great, especially if the area is scanned by a slow moving camera with a pinhole aperture, which insures great depth of field. (If more conventional techniques are used than the number of photos depends on the size of the area to be surveyed and the detail desired.)

If the objective is to identify purely biological objects - presumably vegetation or even animals - then considerations analogous to those mentioned in connection with microscopic surveys are relevant.*

*We have not yet considered the possible biological significance of terrain surveys which do not provide portraits of Martian plants or animals. Terrain studies may be presumed to provide much useful information of indirect biological significance.

7.3.1 A biologically significant object of appropriate scale to be resolved may not exist in the area to be surveyed. The size-frequency assumption suggests that there may be no advantage in decreasing the resolving power of the system in that large plants or animals are proportionately less common. (Evidence of biological activity other than plants or animals themselves would seem to be more readily obtained by chemical means than by means, e.g., of a single visual scan of the vicinity of the lander.)

7.3.2 Photographs may be hard to interpret reliably. Although pictures might appear to reveal plants or animals, the likelihood that they will is a function of the likelihood that Martian plants or animals share visual properties with terrestrial ones. Even if photos appear to reveal biological objects, these effects might be false positives due to terrain features, shadows, and so forth. Ruling out such sources of false positives might be done simply and effectively by comparing photos taken at different times so as to rule out shadows and detect changes in position which might be correlated with temperatures and time of day. Color is an important visual index and comparisons of visual and IR photos might also provide useful data. The value of a visual survey not extensive enough to permit such checks is very questionable in that it might not reveal anything visually interesting and even if it does such photos cannot be reliably interpreted as providing conclusive evidence. The frequency of "false positives" which result when attempts are made to interpret the significance of arbitrary photographs has not been studied directly. However there is much data regarding the effects of context, emotional state and prior knowledge on the perception of non-photographic visual data to suggest that the interpretation problem will be a severe one. Unless experiments which demonstrate the contrary can be conducted, it would be very unwise to suppose that Martian biota will be readily recognized from photographs.

7.4 Preliminary Conclusions

The preceeding examination of factors relevant to the evaluation of visual experiments suggests the following tentative conclusions:

7.4.1 The value and conclusiveness of limited visual surveys, whether macroscopic or microscopic, depend upon the validity of the assumption that recognizable organisms of appropriate size exist in the vicinity of the lander.

7.4.2 The capacity (number of photos) and level of sophistication required in a reliable visual experiment is a function of the visual similarities between terrestrial and Martian organisms. The more closely the photographed orgainsms resemble terrestrial organisms, the less need for large amounts of visual data which will enhance the recognizability of microorganisms or rule out false positives in macrophotos. If Martian life forms satisfy geocentric assumptions, a limited visual reconnaissance experiment may provide useful evidence. The more the environment and life forms of Mars differ from those of Earth, the less likely it is that a batch of photos collected in a predetermined fashion will contain an unambiguous picutre of an organism.

7.4.3 Much relevant experimental data regarding the ability of trained biologists to interpret reliably the kinds of photographs which might be obtained from experiments of this nature can and should be obtained.

8.0 STATISTICAL MEASURES OF LIFE-LIKENESS

One of the problems associated with the design and implementation of a near optimal strategy for the biological exploration of Mars is that of identifying a suitable set of recognition properties upon which decisions regarding the probable presence or absence of life can be based. Desirably, such recognition properties should satisfy the following conditions:

1. They should be properties whose occurrence cannot readily be explained in terms of abiological mechanisms; if possible they should be unique to living systems.
2. They should be general properties which can be expected to characterize all living systems, even the most alien.

One property which has frequently been proposed as the primary characteristic distinguishing living from non-living systems is the high degree of orderliness such systems exhibit as a concommitant of the fact that they are not in thermo-dynamic equilibrium with their environments. It has been frequently observed that such systems contain "information" or "negentropy", which fact is reflected in their metastability and in the observed high degree of structural and functional organization they possess.

This idea of the fundamentally organized nature of living systems is so central to all concepts of the nature of life that it may be supposed that all living systems will exhibit orderliness no matter how alien in origin or structure they may be. The property of orderliness thus satisfies the two conditions listed above. It is not, however, a useful property on which to base decisions regarding the possible presence of life on

Mars because general procedures for dectecting and estimating the amount of order present in systems--on earth or elsewhere--have not been developed. Until such procedures become available and make it possible to test experimentally the claim that living systems are highly ordered--and in this respect differ sharply from non-living systems--this assertion must be dismissed either as an empirically meaningless statement or as a metaphorical, (and not particularly useful) way of describing other properties known to occur in living systems.

During the reporting period just completed the principle investigators initiated a joint study with Dr. J. E. Lovelock of the University of Houston, directed toward exploring the possibilities of generating useable measures of the amount of order exhibited by living systems. Dr. Lovelock has demonstrated that certain hydrocarbons of a crude oil approaching thermo-dynamic equilibrium are highly disordered relative to comparable hydrocarbons in samples of recent biological origin. This difference in orderliness can often be determined by inspection of chromatograms of the corresponding materials. Even when such differences in orderliness are not readily discernable to the naked eye, appropriate comparison of the observed distribution to a suitably defined random one reveals such a difference.* The objective of the joint efforts undertaken with Dr. Lovelock has been to develop a special purpose measure of "information" common two chromatograms, and to devise and test the

---

* The actual techniques employed to effect such a comparison and the results of several analyses will not be described here as they were developed by Dr. Lovelock and Mr. Peter Simmons, also of the University of Houston, and are the subject of a forthcoming paper.

efficacy of alternative procedures for estimating the "amount of order" observed in a single chromatogram. Because relatively few definitive results have been achieved so far, we will confine this report to a presentation of the general orientation of the current work and brief description of the nature of the efforts completed to date.

8.1 Orientation and Approach

A living system may be conveniently viewed as a collection of component chemicals each of which may be described in terms of physical and chemical properties such as molecular weight, free energy of formation, kind and number of its atomic constituents or other relevant descriptive properties. If we know the relative quantities of each component present in the system, then the total system may be described in terms of the resulting distribution of each property employed in characterizing the components, or in terms of the joint distribution of two or more such descriptive properties. Each system is one of many possible configurations of its component chemicals, and each distribution of descriptive properties is similiarly one of many such configurations of descriptive properties. Both the system configuration and the property distribution may be probable or improbable ones relative some reference sets of which each may be said to be a member. To say that the entropy of the system is low relative to some reference class consisting of different chemicals of known relative abundances is equivalent to saying that the system configuration is unlikely to occur as a result of some known or assumed process which selects samples from the reference class. Similiarly, to say that some distribution of descriptive properties appears ordered (and hence indexes the relative unliklihood of the original system configuration) is equivalent to saying that the distribution is also unlikely to occur as the result of some assumed property distribution generating mechanism. In order to determine whether a particular system (or its corresponding distributions of properties) is improbable it is necessary to identify a relevant reference class and to determine the relative probability of occurrence of the system in question in that reference class (together, perhaps, with an estimate of the total number of possible configurations and the relative liklihoods of each). Given an appropriate reference class, an estimate of the number of possible different system configurations and their relative probabilities of occurrence may either be deduced from theories regarding the processes which give rise to the reference class, or may be approximated through empirical observation of a large number of representative members.

Thus the problem of developing a restricted measure of information of the kind desired may be viewed as one of estimating the probability of occurrence of an observed collection of chemicals under the assumption that this collection represents a randomly selected sample of the chemicals in some reference class. Two collections of chemicals can be said to share information to the extent that subsets of each are sufficiently similiar to have been generated by random selection from the same reference class. A subsidiary and more general problem is that of estimating the relative liklihood of a given collection when the parameters of the reference class are not known with certainty.

8.2 Activities to date

The efforts applied to this problem have consisted of a) statistical analysis of the distributions of straight-chain hydrocarbons of varying sources and ages, and b) development of a number of computer implementable models to aid in the theoretical examination of alternative proposed information measures. These efforts are briefly described below:

1. Comparison of distributions of alkanes in samples of biological and abiological origin. Several conventional non-parametric statistical tests have been used generate estimates of the liklihood that two observed alkanes could have been obtained by various kinds of "random" selection procedures from the same reference class. Although these tests differ in the extent to which they reflect known differences in origin of the materials, most of the differences revealed by these tests were in the right direction.

2. Model building. The models which have been defined (but not yet programmed) include:

   a) A stochastic model of a simplified "chemical evolution" procedure which gives rise to a large number of chemical species and thus provides a mechanism for generating arbitrary reference classes defined in terms of relative abundance and persistance of the various species. This model can also be used to effect arbitrary ageing of a given system by degrading its components and gradually driving them to a state of internal equilibrium.

   b) A system generating model which draws samples from a reference class in a variety of prescribed ways, some of which give rise to biased samples.

c) A general purpose model of an idealized gas chromatograph which can be used to stimulate a chromatographic analysis of a simulated material system. This model provides for modifying in a semirandom way the separation and response characteristics of the simulated chromatograph so as to give rise to chromatograms which are realistic in that only a few of the volatile components of the sample yield clear well defined peaks. This corresponds to the case in which only a little of the sample can be analyzed and in which there is as a result great uncertainty regarding the actual parameters of the reference class. (This model has already proved useful in the bandwidth allocation studies described earlier.)